1997

An Analysis of the Construct-Related and Ecological Validity of the Barrow Neurological Institute Screen for Higher Cerebral Functions

Peter J. Wass
University of Windsor

An Analysis of the Construct-Related and Ecological Validity of the Barrow Neurological Institute Screen for Higher Cerebral Functions

by

Peter J. Wass

B. A. Laurentian University, 1989

M. A. University of Windsor, 1991

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the
Requirements for the Degree
of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

1997

854810

APPROVED BY:

Dr. G. Prigatano, External Examiner
Barrow Neurological Institute

Dr. P. Weir
Kinesiology

Dr. W. Balance

Dr. D. Reynolds

Dr. D. Shore, Advisor

Dr. R. Frisch, Chair
Psychology

ABSTRACT

The Barrow Neurological Screen for Higher Cerebral Functions (BNIS) was developed by Prigatano (1991) and his colleagues as a means of providing comprehensive information about higher cerebral functioning for a wide variety of neurological patients with known cerebral impairment. In the development of any psychometric instrument, it is critical to establish its reliability and validity. The purpose of this study is to explore the construct-related and the ecological validity of the BNIS in a patient sample independent of the original patient data gathered by Prigatano (e.g., Prigatano, 1991; Prigatano, 1994). The construct-related validity of the BNIS is explored through its relationship with various neuropsychological measures. Significant positive correlations were established between various neuropsychological measures and subscale items from the BNIS. Multiple regression analyses indicated that 16 percent of the variance in WAIS-R Performance IQ scores was predictable from BNIS Visual Spatial/Visual Problem Solving subscale items, and 35 percent of the variance in WAIS-R Verbal IQ scores was predictable from BNIS Speech and Language subscale items. These relationships provide support for the construct-related validity of the BNIS.

The ecological validity of the BNIS was examined through its relationship with the Functional Independence Measure (Guide for the Uniform Data Set for Medical Rehabilitation, 1993) and its adjunct the Functional Assessment Measure (FIM+FAM) (Hall & Johnston, 1994). Significant positive correlations were found between the BNIS and the FIM+FAM. Multiple regression analyses provided significant solutions for various FIM+FAM Psychosocial-Cognition items as derived from BNIS Subscale items. The results of these analyses provide support for the ecological validity of the BNIS.

ACKNOWLEDGEMENTS

I wish to convey my appreciation to the members of my committee for all their help in the completion of this project, as well as their patience in seeing it through to completion. I would like to thank Dr. Weir for her suggestions regarding the organization and structure of this project. Dr. Reynolds was consistent in his comments that allowed me to focus upon the relevant aspects of this study, as well as his editorial suggestions. Dr. Balance not only provided comments and suggestions that were helpful in the completion of this project, but has also had a significant positive influence on the development of my clinical skills that will continue to influence my interactions with others long after the completion of my dissertation. I wish to also thank Dr. Rourke for the opportunity to be involved in the exciting and challenging career of clinical neuropsychology. Dr. Prigatano's influence over this project was significant primarily through the contributions that he has made to the field of neuropsychology. His ideas and work are and continue to be a considerable inspiration to my work in the field with brain injured patients. Dr. Shore's contributions to this study and my education are especially appreciated. I admire his ability to treat others with a sense of dignity and respect, while maintaining a sense of humor about the human condition.

I would also like to thank all the individuals at Alberta Hospital Ponoka who provided me with the support and encouragement to complete this project. Thanks also to the support staff at the University of Windsor, particularly Barb Zakoor, who assisted me with the completion of all the administrative tasks.

Finally, but most important, I would like to thank all of my family members for their support and encouragement throughout this entire process. My deepest and most sincere gratitude is extended to my parents, Herbert and Dorothy Wass, and my mother-in-law, Catherine Percy. Without their encouragement and help this would not have been possible.

Words cannot begin to express the gratitude that I feel toward my wife for her support, encouragement, and love. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

In recent years there has been growing interest in the development and use of cognitive screening instruments. Many different cognitive and mental status screening instruments have been developed over the past couple of decades. These include the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), the Dementia Rating Scale (DRS; Mattis, 1976), the Cognitive Capacity Screening Examination (CCSE; Jacobs, Bernhard, Delgado, & Strain, 1977), the Short Portable Mental Status Questionnaire (Pfeiffer, 1975), and the Blessed Dementia Scale and Blessed Information - Memory - Concentration Test (Blessed, Tomlinson, & Roth, 1968). Many of these instruments were developed to address the need for a brief and easily administered cognitive screening instrument that would use quantitative information to help identify patients with "organic" impairments admitted to a variety of medical and psychiatric settings.

With increasing recognition that these patients typically require significantly different treatment regimes and exhibit unique prognoses it was important to develop these screening instruments in an attempt to provide the necessary interventions as soon as possible. Ideally, these or any other cognitive screening measures would be included as part of a comprehensive admission procedure and be used to help discriminate those patients who were likely to need a more comprehensive evaluation. Cognitive screening measures could also potentially be utilized to monitor ongoing changes in the patient. In a comprehensive review of the literature of a number of cognitive screening measures, including the ones described above, Wagner, Nayak, and Fink (1995) concluded that at the present time no one instrument is clearly superior to the others and that each has its strengths and weaknesses depending partially on the population to which it is applied.

Full examination of a referred patient or client should include such components as a detailed history of the presenting illness or problem, review of all potential risk factors that may cause or exacerbate the presenting problem, medications and their possible side-effects, full medical evaluation, past psychiatric history, genetic risks determined through family history, careful behavioral observations, and a comprehensive neuro-cognitive examination (Wagner, Nayak, & Fink, 1995). However, because of time constraints and other resource availability factors it is not always possible to obtain a full examination in a timely manner. In these situations screening instruments can be invaluable as a first stage in helping to identify the type of services that any given patient would be most likely to require. Accurate diagnosis of a patient will be crucial to their receiving appropriate interventions and treatment. Therefore, the screening measure should be used only as an initial adjunct instrument to aid in diagnosis until a more comprehensive set of measures can be employed. Also, it would typically not be necessary to put every patient through full physical and psychological examinations. Even if this were possible through the availability of unlimited resources, it can produce unacceptably high false positive rates of identification because of the less than perfect psychometric qualities of any instrument presently available (see discussion below on the predictive power of tests). The consequences for the false positive and false negative identifications will vary depending on the types of problems being identified and the subsequent treatment or lack thereof.

Malloy et al. (1997) encourages the use of cognitive screening instruments for several reasons. The first argument in support of the use of cognitive screening instruments is the likelihood of overlooking cognitive impairment when a formal cognitive screen is not employed.

The second issue is because the use of a brief standardized cognitive assessment with adequate norms can reduce the possibility of misdiagnoses as a result of interpretation bias based on such factors as age or education. The third reason was the cognitive screening measures could contribute to differential diagnoses through the identification of cognitive deficits. A fourth benefit is that screening measures can help delineate the need for further and more detailed assessments, such as a full neuropsychological assessment. A fifth benefit is that the identification of cognitive deficits identified via screening measures may assist in the determination of long-term prognosis. The final reason cited is the suggestion that "formal cognitive evaluation as a measure of treatment outcome may soon be mandated for managed care and public policy analyses" Malloy et al., 1997, p. 190). Malloy et al. (1997) emphasized that an "ideal" cognitive screening instrument should briefly assess all major cognitive domains, including orientation, attention/concentration, language, spatial, memory, and executive functions.

Cognitive Screening Measures

Most of the screening instruments developed to date provide a single total score for all items in the measure. This total score is then used to determine a cutoff point for a given patient population to distinguish between those with or without a given condition (e.g., dementia). Although many clinicians may attempt a quantitative and/or qualitative analysis of the patient's performance on the various tasks, the measures themselves were not specifically designed to provide this breakdown of abilities. These measures briefly assess a variety of cognitive functions, thereby providing a global indicator of cognitive functioning; they were not designed to provide a breakdown of performance based on these various abilities. The Barrow Neurological Institute Screen for Higher Cerebral Functions (BNIS; Prigatano, Amin, & Rosenstein, 1993) has recently been developed. Prigatano and his colleagues specifically set out to develop a measure (i.e., the

BNIS) that provides a break down of cognitive functioning based on commonly identified cognitive abilities (e.g., attention, language, visual spatial, and memory). A similar instrument is the Neurobehavioral Cognitive Status Examination (NCSE; Northern California Behavioral Group, Inc., 1988). The NCSE has been subjected to a broader range of studies conducted by researchers in a variety of clinical settings compared to the BNIS (see [Wass, 1996a] for a review of the literature on the NCSE). This study will focus upon the BNIS in an attempt to expand the available research.

BNI Screen for Higher Cerebral Functions

Prigatano and others developed the BNIS as a means of providing comprehensive information about higher cerebral functioning for a wide variety of neurological patients with known cerebral impairment. Prigatano (1991) argued that existing screening instruments may provide useful quantitative information for patient classification into one nosological category or another (e.g., demented or not), but do not provide adequate information about language, memory, visuospatial, and prefrontal functioning. Prigatano goes on to suggest that in neurological settings a rapid, reliable, and valid screening examination is "needed to aid neurological diagnosis, to monitor the patient's recovery while in the hospital, and to provide useful information to guide both in- and outpatient management" (1991, p. 2).

Prigatano (1991) further suggests that there are four major problems with the available screening measures. The first problem is that there is no systematic evaluation of when a patient is ready to receive an assessment, based upon their level of arousal, basic communication skills, and degree of cooperation. The second problem is that current screens provide only quantitative information, but not the qualitative information about test item performance that can be useful in the diagnosis of neurobehavioral syndromes. The third problem identified by Prigatano (1991) is lack of coverage across the relevant cognitive domains, including language, visuospatial, learning and memory, impulse control, interpretation and control of affect for patients with focal, bilateral, diffuse, or lateralizing impairments. A final concern is the inadequate attention in the screening measures to the patient's self-perception or awareness of cognitive functioning. Prigatano also provides some general guidelines for developing any screening measure, which includes ease of administration and scoring, brevity, reliability (e.g., same results by different examiners), and validity (e.g., identifies areas of neuropsychological impairment). Prigatano (1991) further suggests that a screening measure should be capable of monitoring clinically significant changes, aid in clinical management decisions, and assist in determining the prognosis for a patient. Prigatano indicates that the development of the BNIS proceeded with all of these factors in mind.

The BNIS manual states that only appropriately trained individuals should be allowed to administer the screening test. The BNIS manual specifically states that:

> the examiner should be an experienced clinician with training in behavioral neurology and neuropsychology. Qualified individuals typically include neurologists and psychologists with special training in clinical neuropsychological assessment. Other experienced clinicians (i.e., physiatrists, speech and language pathologists) might also administer this test depending on their experience and expertise. (Prigatano, Amin, & Rosenstein, 1993, p. 17).

This level of training requirement for the BNIS is suggested because the BNIS "was developed to provide a standardized method of enhancing the clinician's initial observations" and, further, that no test can substitute for clinical judgment and experience (Prigatano, Amin, & Rosenstein, 1993, p. 18). At the present time there is no clear indication from published research studies that the BNIS provides superior psychometric results or clinical information because of this administration requirement. Therefore, the decision about who should administer the BNIS would be based upon the professional judgment of the clinician who is responsible for the patient's well-being and treatment. Clearly, the level of interpretation regarding the patient's performance on each task will be dependent upon the level of expertise and training of the examiner. An examiner with minimal test administration qualifications is likely to miss many relevant qualitative details about the patient's performance.

The first element of the BNIS involves the determination of whether or not the patient is capable of reliably responding to the screening items (Prigatano, Amin, & Rosenstein, 1993a). This includes a ranking of their basic level of consciousness/alertness, their basic language functioning, and their cooperation with the examiner. Each of the three items are scored on a three point scale, with directions in the manual indicating that testing should proceed only if a score of at least 2 is achieved on each item. The subsequent 35 test items are grouped into seven subscales (see Appendix A and Appendix B). These include A) Speech and Language Functions, B) Orientation, C) Attention/Concentration, D) Visuospatial and Visual Problem Solving, E) Memory, F) Affect, and G) Awareness versus Performance. The only materials required for test administration are the nineteen laminated 3 x 5 cards, paper, and pencil. Prigatano (1991) indicated that the BNIS took approximately 15 minutes to administer to the majority of their subjects, with no patient taking longer than 30 minutes. Appendix C provides information on the scoring and subscale test items.

In Prigatano's 1991 article on the development of the BNIS he presented the results from two initial validation studies. In the first study, 40 patients from the Barrow Neurological Institute and 20 controls (primarily hospital personnel and therapists) were tested. Traumatic brain injury (TBI) was the most common diagnosis (60 %) in the patient group. It was indicated that most of the patients had diffuse injuries, although specific information regarding the method of diagnosis (e.g., brain scans, medical work-up) was not provided. The control group was not significantly different in age, but the female-to-male ratio was higher and the education level was significantly higher by about 2 years ($t = 2.1$, $df = 52$, $p = .04$). Normative data from this first study was provided. The control group and the patient group were divided into two age groups; younger than or older than 55 years old. Prigatano used this age cutoff based on the argument that "due to normal aging factors, individuals older than 55 years can be misdiagnosed as having cerebral impairment by various neuropsychological tests" (1991, p. 5). The patient group is further divided into patients whose lesions had been sustained within the past 6 months, and those with lesions existing longer than 6 months.

The younger than 55 years old control group had an average score of 48.3 out of fifty. Therefore, a cutoff of 48 was utilized to determine the presence of true positives and true negatives in the entire patient group (Prigatano, 1991). It was reported that all patients were correctly identified using this cutoff, and 87.5 percent of the control group. The subsequent Sensitivity of the test was reported to be 94.7 percent, and the Specificity was 87.5 percent for the test in the younger age group.

In the second study from Prigatano's 1991 article, a group of 50 patients and 22 controls were tested in an attempt to cross-validate the first study's findings. Prigatano indicated that "the scoring of certain items was modified slightly", and "a few new test items were added and others were dropped" (1991, p. 5-6). However, no further information was provided about these changes. The demographic composition of the patient and control group was similar to the groups in the first study (i.e., age similar, education and female-to-male ratio higher in the control group). The patient group was once again predominantly TBI patients (60 percent). The cutoff score was kept at 48 of a possible 50 points. As with the first study group, the acute patients did more poorly than the chronic patients. The Sensitivity for the entire age range in the second study was 96 percent and the Specificity was 64 percent. Prigatano (1991) argued that a cutoff point approach is more likely to misclassify older individuals as brain damaged. Therefore, it was suggested that the inclusion of the entire age range in the second study resulted in lower Specificity from that obtained in the original sample.

The overall Sensitivity of the BNIS was certainly acceptable based upon these initial studies, but the Specificity values were low when a large age range of subjects was analyzed. Prigatano (1991) further explored the Delayed Recall and Awareness tasks to improve the Specificity. He argued that memory complaints are common in both the brain injured and the elderly. However, elderly individuals with no identifiable brain injury can do poorly on these measures as a result of the normal aging process or depression, among other things, as well as brain impairment. This analysis was performed on a total of 111 patients and the 42 control subjects to see if these two tasks would help improve the discrimination between patients and control subjects. On the Delayed Recall task, 4 of 42 controls were unable to recall all three words, while 72 out of 111 patients were unable to recall the three words. Therefore, the Delayed Recall task's Sensitivity was only 65 percent, but its Specificity was 90 percent (false positive ratio = 5 %; false negative ratio = 51 %). On the Awareness task, 3 control subjects did not predict their recall accurately, while 73 patients were unable to predict their recall. This resulted in 66 percent Sensitivity and 93 percent Specificity for the Awareness task (false positive ratio = 4 %; false negative ratio = 49 %). This would suggest that Specificity could be improved through analysis of these two tasks for any given patient, but that the Sensitivity is clearly higher with the overall cutoff point system.

Both the control group and the patient group from the second study were tested with the MMSE to assess the construct validity of the BNIS. As described by the Committee to Develop Standards for Educational and Psychological Testing (1985), a strong relationship with other measures that are purported to identify the construct in question provides evidence for the construct validity of the test instrument. Prigatano stated that "a positive and significant correlation was anticipated based on the assumption that both tests were sensitive to brain dysfunction. However, because the BNI Screen attempts to assess abilities not sampled by the MMSE, the correlation of these two tests was anticipated to be between .50 and .75, reflecting a common variance between 25% and 60%" (1991, p. 6). He argued that if the correlation between the BNIS and the MMSE was .85 or higher then the BNIS would essentially be measuring nothing new. Based upon the results of the patient and control group study a correlation of .71 was reported between the MMSE and the BNIS and was interpreted as providing support for the construct validity of the BNIS.

Rosenstein, Prigatano, and Amin (1992) conducted studies on the test-retest and inter-rater reliability of the BNIS. The test-retest reliability is particularly important considering Prigatano's (1991) intention that the BNIS be a useful measure of change over time. High test-retest reliability would help ensure that changes in performance over time in a patient would reflect real changes in their functioning. A total of 13 female and 19 male patients from the initial validation sample were retested an average of 2.9 days (S.D. = 2.9; range 0.08 - 9.0) after the first administration. Twenty-one of the patients were retested by the same examiner, with the remaining 11 patients retested by a second examiner. The only change in test procedures between testing was the use of two different word lists in the Delayed Recall task.

It was reported that the test-retest reliability coefficient for the entire sample was $r[30] = .940$ ($p < .001$). The test-retest correlation for the 21 patients who had the same examiner was

.970, and for the other 11 patients with different examiners was .911. The correlation between the test interval and changes in score was not significant. The test-retest reliability for the entire sample on each of the subscales was as follows: Speech and Language, .931; Orientation, .449; Attention/Concentration, .601; Visuospatial, .836; Memory, .809; Affect, .616; and, Awareness, .309. The test-retest correlation between the two sets of words used in the Delayed Recall task was significant at .749. Based upon the criteria for clinical significance of reliability coefficients established by Cicchetti and Sparrow (1981), the Speech and Language, Visuospatial, and Memory subscales were within the excellent range. The Attention/Concentration and the Affect subscales were within the good range. The Orientation subscale fell in the fair range, and the Awareness subscale fell within the poor range. Rosenstein, Prigatano, and Amin (1992) suggest that practice effects may have played a substantial role in regard to the low test-retest reliability of the Orientation and the Awareness subscales. In a retrospective analysis of the Orientation responses, 11 of the 32 patients failed on the first test, while 7 of those 11 patients subsequently gave the correct response on the second testing. Of the 21 patients who received full marks on the first testing, only 3 failed on the second testing. A similar pattern was reported for the Awareness subscale. Twenty-three of the 32 patients failed this item on the first testing, although 10 of those 23 patients correctly guessed their performance on the second testing.

To test the inter-rater reliability of the BNIS a group of 10 patients was tested by one examiner as another examiner sat in the room and judged the patient's score on each task. On four tasks the subject is given a second chance to respond if they are incorrect on the first trial. For those four items both trials were administered regardless of the initial answer so that the second examiner would not know how the other examiner scored the patient. The overall correlation between the two examiners was significant at .998 ($p < .001$). The percent of agreement between the two examiners on each of the tasks was perfect for all but 5 tasks. The percent of agreement for these five tasks was 90 % on basic language, 80 % on paraphasia, 90 % on reading, 90 % on writing, and 70 % on affect expression. Rosenstein, Prigatano, and Amin (1992) indicated that the directions for scoring the items on paraphasia and affect expression were revised as a result of these findings.

Regarding the use of percent agreement, rather than a kappa coefficient, it was argued that kappa is limited by the actual prevalence of the characteristic being rated, with kappa approaching zero as the prevalence approaches zero or one (Rosenstein, Prigatano, and Amin, 1992). Therefore, it was suggested that a kappa could not be calculated since even seriously impaired patients will typically pass some tasks on the BNIS. Rosenstein, Prigatano, and Amin (1992) indicate that those items that are rarely failed (e.g., visual object recognition) are still important because they measure functions that are infrequently impaired but, when present, are significant for clinical diagnosis and patient management.

Rosenstein, Prigatano, and Amin (1992) concluded that the above described reliability studies provided evidence that the BNIS total score was reliable across administrations and between administrators. Prigatano and his colleagues then proceeded to explore the validity of the BNIS. The construct validity of the BNIS was assessed through comparison of BNIS scores with MMSE scores for 105 patients and 49 control subjects (Prigatano, Amin, & Rosenstein, 1993b). The order of presentation for the BNIS and the MMSE was counterbalanced across subjects. Mean scores for the control and patient groups differed significantly on the BNIS and the MMSE. The reported correlation between the BNIS and the MMSE was $r[103] = 0.808$, $p < .001$, producing a shared common variance of 65.3 percent. This was higher than the 0.71 correlation coefficient found between the BNIS and the MMSE reported in the initial validation study of the

BNIS (Prigatano, 1991). Prigatano, Amin, and Rosenstein, (1993b) also reported that patient or control group membership explained 30 percent of the total variance of the BNIS scores ($R^2 =$ 0.298) but only 13 percent of the total variance on the MMSE score ($R^2 = .134$). Based on this finding they conclude that the BNIS "captures a greater degree of variance attributed to brain damage or dysfunction than the MMSE" (Prigatano, Amin, & Rosenstein, 1993b, p. 4).

The second study described by Prigatano, Amin, and Rosenstein (1993b) detailed an attempt to establish the concurrent validity of the BNIS. They analyzed how well the BNIS was able to discriminate patients with brain impairment from subjects without any recognized brain impairments. The Committee to Develop Standards for Educational and Psychological Testing (1985) indicate that the study design best suited to gathering evidence for establishing criterion-related validity in diagnostic tests is through the concurrent administration of the criterion and prediction measure. A total of 225 patients with known brain impairments and 102 subjects with no history of brain impairment or psychiatric disability were compared. Once again, Prigatano and his associates do not explain how the patients were diagnosed with brain impairment. The average score obtained by the control subjects was slightly higher than 46, leading Prigatano, Amin, and Rosenstein (1993b) to use a cutoff score of 47. This was one point lower than the cutoff score used in the Prigatano (1991) initial validation study. As with the initial validation study (Prigatano, 1991), all of the subjects were divided into two groups based on age (i.e., < 55 years old, and ≥ 55 years old). The patients were further divided into those whose lesions were sustained less than 6 months earlier compared to those with lesions sustained longer than 6 months ago.

When comparing all patients and all controls, the overall Sensitivity was an acceptable 92 percent with a false positive ratio of 18 percent, and the Specificity was a low 56 percent with a false negative ratio of 24 percent. An analysis of the subjects less than 55 years old produced a Sensitivity ratio of 90 percent with a false positive of 13 percent, and Specificity of 70 percent with a false negative of 24 percent. As can be seen, the values for the younger age group are generally similar to the overall test result ratios. Analysis of the older than 55 years age group resulted in a Sensitivity of 98 percent with a false positive of 28 percent, and a Specificity of 88 percent with a false negative of 75 percent. Although the Sensitivity and Specificity of the BNIS is higher in this older age group, it comes at the expense of a considerably increased risk of false positives and false negatives. Utilizing the same rationale presented in the initial validation study (Prigatano, 1991), Prigatano, Amin, and Rosenstein (1993b) further analyzed the Delayed Recall and the Awareness subscales. When the patients and controls were classified as brain impaired or not based upon these subscale scores, the Sensitivity dropped to 67 and 68 percent, and the Specificity increased to 93 and 91 percent for the Delayed Recall and Awareness subscales, respectively. Based upon the results of these analyses, Prigatano, Amin, and Rosenstein (1993b) indicate that the total BNIS score provides a good estimate of Sensitivity and a more detailed analysis of the Memory performance serves to increase the Specificity.

The third part in the series of validity studies conducted by Prigatano, Amin, and Rosenstein (1993b) was an analysis of the BNIS's ability to distinguish patients with right or left hemisphere damage. Adequate performance on the Speech and Language tasks was assumed to indicate a normally functioning left hemisphere, while the Visual Spatial and Visual Problem Solving tasks, as well as the Affect tasks, requires an intact right hemisphere. Fourteen patients with documented (not specified how) lesions of the left cerebral hemisphere, and 14 patients with documented lesions of the right cerebral hemisphere were compared to 72 control subjects that were matched for age and education level. Prigatano, Amin, and Rosenstein (1993b) indicated that all of the subscale performance predictions were supported by the test results. In other words, the group of patients with right hemisphere lesions did significantly worse on the Visual Spatial and the Affective subscales compared to the other patient group. The patients with left hemisphere lesions did significantly worse on the Speech and Language subscale. Also, there were no significant differences between the two patient groups on any of the other four subscales. The control group performed significantly better on all subscales than either patient group.

The draft version of the manual for the BNIS provides further information regarding the psychometric properties of the measure (Prigatano, Amin, & Rosenstein, 1993a). At the time of the writing of this paper a final version of the manual had not been released. The manual provides normative data for a total of 255 patients and 102 control subjects based upon the most recent administration version of the BNIS. Within the patient group the mean age was 41.3 years (S.D. = 19.1), the average level of education was 12.6 years (S.D. = 2.9), and was composed of 162 males and 93 females. The control group had a mean age of 40.0 years (S.D. = 18.9), average level of education of 13.8 years (S.D. = 2.3), and was composed of 40 males and 62 females. Education was reported to be significantly higher in the control group. The average BNIS score for the patient group was 36.34 (S.D. = 7.6), and the control group had an average BNIS score of 46.36 (S.D. = 2.9). In the control subjects who were younger than 40, the BNIS total score would typically fall above the 47 point cutoff. Scores on the BNIS exhibited a modest decline with increasing age from the mid-50's through to old age. Analysis of the correlation between age and total BNIS scores indicated a much stronger relationship for the control group ($r = -0.67$) than for the patient group ($r = -0.28$). The correlation between age and BNIS scores in the control group remained significant at -0.53 when education was partialed out. There was a similar correlation pattern between the BNIS scores and education for the control group ($r = 0.50$) and the patient group (r = 0.22). When the effects of age were partialed out for the control group the correlation between education and BNIS scores dropped to a non-significant 0.16.

When the larger samples from the manual are used with a cutoff of 47 points Sensitivity of the BNIS was 93 percent, with a false positive ratio of 16 percent. The Specificity was reported to be only 50 percent, with a false negative ratio of 24 percent. Once again, Prigatano and his colleagues recommended using performance on the Delayed Recall task as a way to improve the Specificity. The Specificity increased to 93 percent when Delayed Recall was used to discriminate patients from controls, but the false negative ratio increased substantially to 46 percent. This method also produced a significantly lower Sensitivity of 68 percent, with a false positive ratio of only 4 percent.

A study by Prigatano and Smason (1994) compared the performance of 20 Alzheimer's patients with 20 age and sex matched normal control subjects. The diagnosis of Alzheimer's disease was based upon the DSM-III-R criteria for dementia of the Alzheimer's type. Mean age for the patients was 73.9 years and 72.3 years for the control subjects. The Alzheimer's disease patients were also administered the MMSE. The average score on the BNIS for the patients was 29.15 (S.D. = 6.22), and the control subjects had an average BNIS score of 44.2 (S.D. = 3.64) out of a possible 50 points. This difference was highly significant. The Alzheimer's patients performed significantly worse than the controls on every subscale of the BNIS. Prigatano and Smason (1994) also reported on the Sensitivity and Specificity of the BNIS based on the results of this study group. With a BNIS score cutoff point of 40, all of the Alzheimer's patients were correctly classified, and 18 out of twenty controls were correctly classified. The Sensitivity was 100 percent and the Specificity was 90 percent, with a false negative ratio of 9 percent. The reason for a cutoff point of 40 was not clearly stipulated by Prigatano and Smason (1994). The expected mean value for normal subjects between 60 and 70 years old was a score of 43 points (Prigatano & Smason, 1994). When this cutoff point was used the "estimate of sensitivity was 100 % and the estimate of specificity was 80 %" (Prigatano & Smason, 1994, p. 32). The correlation between the BNIS and the MMSE was reported to be statistically significant at 0.546, which accounted for about 30 percent of the shared variance.. Prigatano and Smason (1994) concluded that this study provided encouraging data on the utility of the BNIS in screening patients with suspected Alzheimer's disease.

An issue that is explored in the draft manual is the correlation between various neuropsychological measures, the BNIS subscales, and the total BNIS score for 14 control subjects and 102 of the patients from the normative data sample. The neuropsychological measures employed are the Wechsler Memory Scale - Revised (WMS-R; Wechsler, 1987) Logical Memory I & II and Visual Reproduction I & II subtests; the Trail Making Test A and B; and the Wechsler Adult Intelligence Scale - Revised (WAIS-R; Wechsler, 1981) Digit Symbol, Picture Completion, Block Design, and Information subtests. All measures were at least moderately correlated except for the correlation between the BNIS Orientation subscale and Trails B. Correlations between the neuropsychological measures and the BNIS total score were particularly strong. Correlations between the BNIS Visuospatial subscale and the WAIS-R Block Design, Picture Completion, and Digit Symbol subtests were quite strong (i.e., .652, .631, & .613, respectively). Also, the correlations between the BNIS Memory subscale and the four WMS-R subtests were quite strong, ranging from 0.729 to 0.795. The correlations between the BNIS subscales and the neuropsychological measures were presented as further evidence supporting the construct validity of the BNIS.

The issue of veridicality for the BNIS had received some attention by comparing the BNIS scores to the Patient Competency Rating Scale scores (PCRS; Prigatano, Altman, & O'Brien, 1990). Veridicality is a measure of ecological validity that assesses the "extent to which test results reflect or can predict phenomena in the open environment or 'real world'" (Franzen & Wilhelm, 1996, p. 93). The study by Prigatano and Leathem (1993) compared the performance of 20 New Zealand Maoris and 20 Non-Maoris on the BNIS. All except two subjects were classified as having a severe traumatic brain injury, defined by a loss of consciousness for longer than 24 hours and physician classification. The other two subjects suffered a moderate TBI, as defined by loss of consciousness for at least 1 to 6 hours. The two groups of subjects did not differ in age, education, sex, or time since brain injury. Each patient and their primary caregiver independently rated them on their ability to perform a number of functional daily tasks using the PCRS. The PCRS is composed of 30 questions with a 5 point rating of how well the individual is able to complete a number of activities of daily living and psychosocial functioning. The BNIS scores were then compared with the PCRS scores to explore the relationships between perceived ratings of competency and neuropsychological screening results.

The two groups did not differ significantly in their total BNIS scores. Scores ranged from 20 (severe impairment) to 48 (mild to no impairment), with the average level of BNIS performance falling in the moderately impaired range. The only subscale that was significantly more impaired in the Maori group, compared to the Non-Maori group, was Digits Forward. A number of the Speech and Language items were generally weaker in the Maori group, even though English was the primary language for 18 out of 20 Maori subjects. There was very little relationship between the *self-ratings* of behavioral competency and BNIS performance in both groups. On the other hand, overall *caregiver ratings* of competency were significantly related to the patient's BNIS scores in both groups. These findings provide preliminary support for the veridicality of the BNIS. Further study of this issue needs to be pursued using other functional assessment instruments that are completed by professionals involved in the individual's treatment and care.

In conclusion, the present research on the BNIS supports its efficacy as a cognitive screening measure. However, as a result of the limited range of settings within which research on the psychometric properties has been conducted to date, it is difficult to make generalized statements about its utility outside of the presented validation sample. Additional studies regarding the predictive power of the BNIS in different populations will be required. The BNIS might be more appropriate as a "second-stage" evaluation measure, as described by Fields, Fulop, Sachs, Strain, and Fillit (1992). In a comparison of the NCSE and the MMSE, Fields et al. (1992) concluded that the MMSE is probably a better measure as a "first-stage" cognitive screen for the elderly since it takes much less time to administer, is acceptable to more patients, and correlates well with psychiatric diagnoses. They suggest that the NCSE is probably better used as a "second-stage" screen in order to provide a more fine-grained analysis of cognitive abilities when the MMSE or psychiatric interview finds more "borderline" performances. The similarities between the NCSE and the BNIS in their subscale profile development and psychometric properties (Wass, 1996b) make it likely that the BNIS would also be better utilized as a "second-stage" screening instrument.

When used in this way, the BNIS could provide potentially useful information about the patient's cognitive functioning in a number of areas until a full neuropsychological evaluation could be completed. The information obtained from the BNIS could help determine the focus of treatment, aid in the diagnostic process, and monitor progress through treatment and/or spontaneous recovery. The BNIS has also been found to be related to care-giver ratings of functioning and competency, thereby providing preliminary evidence in support of its veridicality. There is a clear need for further research in a variety of clinical settings using different measures against which to compare the performance of the BNIS.

<u>BNIS Summary</u>

Prigatano and his colleagues attempted to develop the BNIS as an instrument that could provide a brief analysis of an individual's cognitive functioning for those individuals who may require further testing. The BNIS test results may aid in the initial diagnosis, treatment decisions, and monitoring of a patient's progress. It is different from most other cognitive screening measures because it provides subscale summaries of different cognitive abilities along with a global measure of cognitive functioning, rather than focusing exclusively upon a global measure of cognitive impairment.

Insert Table 1. Here

Table 1 presents a summary of the predictive power for the BNIS based on the literature reviewed in this paper. At the present time this is limited by the lack of research from a variety of settings. All of the predictive power data are derived from two validation sample groups at the setting where the BNIS was developed. This data is from two different samples: the initial validation sample (N=136) and controls (N=52), and the expanded validation sample (N=253) and controls (N=102). The expanded validation sample produced modest improvements in almost all predictive power measures. The ratio of patients to controls remained constant at approximately

72 percent across both validation samples. As described by Baldessarini, Finklestein, and Arana (1983), the PPP rate declines sharply as the prevalence of a disorder declines. In the BNIS studies reviewed, the prevalence rate remained constant. Therefore, utilizing the BNIS in settings that have a different ratio of cognitively impaired to cognitively unimpaired patients (i.e., either higher or lower than the validation samples) would result in different PPP and NPP rates. Sensitivity (i.e., true positive rate) was typically very high to excellent (89 % and 93 %), while the specificity (true negative rate) was considerably weaker (50 % and 56 %). Utilizing the Delayed Recall and Awareness tasks resulted in an improved chance of correctly identifying those subjects who were not cognitively impaired. Under this condition the specificity increased from the 50 percent range to the low 90 percent range, although the sensitivity fell to the high 60 percent range. The probability of the total BNIS score correctly identifying a subject as being impaired (i.e., PPP rate) was in the mid 80 percent range. A corresponding increase to the mid-90 percent range in the PPP was found when analyzing the Delayed Recall and Awareness tasks. The probability of the BNIS correctly identifying a subject as being not impaired (i.e., NPP rate) increased from 63 percent in the first validation sample to 76 percent in the expanded validation sample.

The results of these studies indicate that the BNIS is highly sensitive to the presence of cognitive impairment, but that it carries a high chance of identifying a non-impaired individual as cognitively impaired. Prigatano, Amin, and Rosenstein (1993) suggest performing a further evaluation of the subject's performance on Delayed Recall and Awareness to improve the likelihood of correctly classifying an individual as impaired or normal. The predictive power analyses of these two tasks indicated that failure on these tasks reliably classifies an individual as cognitively impaired. It would be necessary to determine if the BNIS will meet the needs of other clinics and their patient populations since the existing literature was based upon data gathered in the same clinic. Therefore, conducting exploratory research would be strongly recommended to determine whether or not the BNIS would meet the needs of other clinics and their patient populations.

One factor that will determine the utility of the BNIS in any setting is the inherent risk in false positive and/or false negative identifications. As this will vary from one setting to another and one patient population to another, it will depend upon the clinical and scientific judgment of the professionals to determine the utility of the BNIS for that patient population and clinical setting. In general, the BNIS would be useful in a setting that requires a highly sensitive measure of cognitive impairment. However, it also exhibits a high likelihood of incorrectly classifying an individual as impaired when other index measures or diagnostic methods do not support that conclusion. The risk involved in this misclassification will depend upon the setting in which it is being applied. If the risk of failing to correctly identify cognitive impairment is greater than incorrectly classifying a normal individual then the use of this measure could be supported.

The present literature available on the BNIS provides promising data on its psychometric qualities. However, it is unlikely that the BNIS would be significantly better than some of the other existing screening measures (e.g., MMSE) in regard to acting as a "first stage" global cognitive screen. The strength of the BNIS resides in its ability to provide a more fine-grained analysis of possible cognitive weaknesses in a variety of cognitive ability domains. This information should prove useful when attempting to develop intervention and treatment strategies for a patient, as well as monitoring their performance across time. Therefore, the BNIS should prove to be a potentially useful addition to clinical interventions and research programs.

The development of the BNIS attempted to provide measures that were sensitive to frontal lobe impairment (i.e., the Awareness subscale). The BNIS also requires the clinician to estimate the individual's premorbid level of cognitive functioning. Further research will be required to evaluate the extent to which the BNIS meets the goal of assessing frontal lobe impairment. Some support presently exists for the BNIS subscale structure based upon the correlation between various neuropsychological measures and the BNIS subscales. However, this finding has not been duplicated in a setting outside of the institution where it was developed. In conclusion, findings from the BNIS should be supplemented by other measures of neuro-cognitive functioning until further studies can provide greater support for its psychometric properties.

Reliability and Validity

An important and critical factor in the development of any test instrument used in the field of psychology is the establishment of its reliability and validity (Committee to Develop Standards for Educational and Psychological Testing, 1985). Reliability has been described as "the consistency of scores obtained by the same persons when reexamined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions" (Anastasi, 1988, p. 109), and is said to refer to "the degree to which test scores are free from error measurement" (Committee to Develop Standards for Educational and Psychological Testing, 1985, p. 19). Reliability is measured through the use of reliability coefficients; a generic term used to describe the statistical procedures and inferences used to identify sources of error variance. Analysis of reliability will often include inter-rater, test-retest, and alternate form analyses when possible.

Validity of a test is described by Anastasi as concerning "*what* the test measures and *how well* it does so" (1988, p. 139), and is grouped under the three principal categories of content-related, criterion-related, and construct-related. As indicated by the Committee to Develop Standards for Educational and Psychological Testing, validity refers to the extent or degree to which the actual results from the test support the inferences made from the test scores. In other words, the concept of validity "refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from the test scores" (1985, p.9). Content-related validity refers to the extent to which the test items effectively represent some pre-defined domain. Expert judgments about test content and item selection are typical within this type of validity analysis. Criterion-related validity refers to the relation between test scores and some other commonly accepted outcome criteria, and indicates the effectiveness of a test to predict an individual's performance in day to day situations outside the test setting (Anastasi, 1988). Construct-related validity is said to focus "primarily on the test score as a measure of the psychological characteristic of interest", which could include such psychological constructs as reasoning abilities, memory, or visual spatial abilities (Committee to Develop Standards for Educational and Psychological Testing, 1985, p. 9). In general, the information obtained and the inferences made from any test will be limited by the reliability and validity of that measure.

Ecological Validity of Neuropsychological Testing

Neuropsychologists are becoming more involved in the active application of treatment, as well as in legal cases, where the actual diagnosis of impairment is only the first step. A question that is being asked with increasing frequency is, 'what are the functional implications of this pattern of cognitive strengths and weaknesses?' This means that neuropsychologists are being asked to make predictive statements about how well an individual will be able to perform a variety of daily personal care, household, and work-related activities based upon the scores obtained in the neuropsychological assessment. Therefore, an important element in the psychometric description of psychological tests would be their ecological validity. Franzen and Wilhelm (1996) indicate that the term 'ecological validity' of neuropsychological tests refers primarily to the relationship between aspects of assessment tasks and everyday behaviors and tasks. Aspects of assessment tasks could include the actual test results or the underlying constructs (e.g., visual spatial ability) governing the development of a task in a test. Franzen and Wilhelm propose that there are two general aspects of ecological validity, especially in reference to neuropsychological testing.

> The first is verisimilitude, or the similarity of the data collection method to tasks and skills required in the free and open environment. The second is veridicality, or the extent to which test results reflect or can predict phenomena in the open environment or 'real world' (1996, p. 93).

It is apparent from these definitions, that verisimilitude is most relevant during the design phase of a neuropsychological test instrument. Once the instrument has been designed veridicality becomes more critical. Evidence would be required to show that everyday behaviors and task performance could be predicted by the assessment instrument results. Goldstein indicated that veridicality refers to "empirical validity based upon studies of correlation between test scores and measures of relevant behaviors" (1996, p. 78).

Ecological validity is likely to overlap other types of validity in various situations. For example, ecological validity could be seen to overlap with predictive validity when attempting to predict success or failure from a test score, or with face validity when attempting to develop a test based upon verisimilitude. With ecological validity the primary focus should be on the relationship of test scores to everyday behaviors. The determination of face validity and predictive validity could be inferred from the same measure or measures, even though they were not developed specifically to assess the individual's everyday behaviors (e.g., a diagnostic test for brain localization).

Franzen and Wilhelm (1996) suggest that pursuing the issue of an instrument's verisimilitude would require an examination of the theoretical relation between the abilities required to perform that test (e.g., attention and visual-spatial skills) and the corresponding functional behavior predicted by the test instrument. Goldstein (1996) suggested that achieving verisimilitude would be a practical impossibility because even if a test or test item could be developed that perfectly duplicated the behavior in question, the assessment would not duplicate the different settings and behaviors that would be typical and relevant to the individual being assessed. The very process of testing the individual places him or her within a contrived situation. On the other hand Goldstein (1996) indicated that there is considerably more potential for determining the veridicality of different instruments since the relationship between a given test item or total test performance and real-life behaviors or functions can be documented and empirically analyzed. Determining the veridicality of the instrument can be accomplished regardless of the verisimilitude of the test item(s).

Determining the veridicality of a test instrument could involve an examination of the relationship between the score(s) on that measure and the score(s) on another measure that has a documented relationship with behavior in day-to-day situations. For example, if scores on a measure of functional ability are shown to predict success in daily living skills after a rehabilitation program, then the relationship between the neuropsychological measure and the measure of functional ability could provide information about the ability of the neuropsychological measure to predict success in daily living skills. However, Franzen and Wilhelm (1996) point out that at least part of the shared correlation between these two measures could be for reasons other than the shared variance of the target behavior. The relationship between the neuropsychological test and the measure of functional ability is likely to be stronger in situations where the measure of the target behavior involves analysis of the behavior in the open environment. This could involve self-reports, reports by others, formal behavioral observations in a natural setting, or questionnaires (Franzen & Wilhelm, 1996).

Franzen and Wilhelm (1996) identify three methods of obtaining ecologically relevant information about an individual's behavior. The first is to develop test measures designed to elicit the behavior of interest. Although it *may* be possible to develop a limited number of measures that would fit this description, some criticize these attempts and suggest that it would not be possible to develop truly realistic measures as a result of the contrived nature of assessment (e.g., see above; Goldstein, 1996). The second would be the direct observation of the behavior in a naturalistic setting, or observation of an individual attempting to perform some standard everyday tasks (e.g., shopping for food and preparing a meal). Once again, when the individual is being observed, it is highly probable that their behavior would not be exactly the same as when they are performing a task unobserved. Also, natural setting observations are not feasible in most situations since the resources required would be quite extensive. The third method identified, would be the use of functional assessment instruments that inquire about the individual's ability to perform various tasks. This can be achieved through questionnaires completed by family members and other collateral sources, or through rating scales completed by various professionals involved with the individual who have knowledge of the relevant aspects of the individual's behavior.

Functional assessment instruments appear to hold the greatest potential since they can usually be completed relatively quickly, can be standardized, and can provide a normative base of information. Franzen and Wilhelm (1996) identify five criticisms leveled at the use of functional assessment instruments. First, they point out that most of these instruments are subjectively scored. They suggest that although this is not necessarily a significant drawback, it is critical to establish the inter-rater reliability of the measure to assess the adequacy of the scoring instructions. The second criticism is that the items to be scored are frequently difficult to judge because the item encompasses a broad range of behaviors potentially required to perform that task (e.g., maintaining personal hygiene). Failure or impairment in any one functional component of that item can result in a failure to effectively complete that task. A third criticism identified by Franzen and Wilhelm (1996) is that the scale properties of many of the functional assessment instruments is unknown. Psychometric evaluation (e.g., reliability) of the instrument cannot be established until the scalar properties of the instrument are known (i.e., are the scale items nominal, ordinal, interval, or ratio). For example, many of the instruments combine many or all of the items into summary scores. Without knowledge of the scalar properties of the various items, this could result in erroneous conclusions since some items may be ordinal in nature, while others within the same instrument could be classified as interval scaling. Analyses should be conducted to determine the scalar properties of any given instrument.

The fourth criticism involves the establishment of appropriate criteria for whether or not the individual can perform the specified task outside of the assessment situation. In other words, will the individual be able to generalize the component tasks to the various day-to-day situations that may arise? This criticism is related to Goldstein's (1996) contention that a test situation task could not be truly "real" in that it could not be directly assessed without the assessment process itself acting as a significant confound. The fifth criticism is related to the previous concern, in that there is a lack of validation studies relating the instrument measures to actual behaviors in the natural environment. This means that the tasks were chosen for the measure based upon what was assumed to be relevant criteria for the individual's daily functioning (i.e., verisimilitude), but the actual relationship between the given tasks and daily functioning still have to be determined (i.e., veridicality). Franzen and Wilhelm (1996) suggest that this is a greater problem with instruments that rely upon the reports of others (e.g., family members), but is still a concern with instruments that rely upon more direct observations of the target behaviors (e.g., rating scales completed by professionals based upon contact with and observation of the individual).

Predictive Power

An important consideration in the use of a cognitive screening measure is the predictive power of the test in various settings and with different patient populations. This refers to the ability of the test to accurately identify those patients who actually have a given condition as opposed to those who do not as determined by another standard or index. Comparison of the test results is ideally made with some well-established and commonly accepted "gold standard". Two commonly presented values are the test's sensitivity and its specificity, which were terms introduced by Yerushalmy (1947; as cited in Baldessarini, Finklestein, & Arana, 1983). Appendix D presents a table that is commonly used to represent the placement of the subjects or patients according to the "gold standard" or index measure and the subsequent findings from the test results. The first vertical column in this table includes those subjects that have been identified as having the condition in question according to the index measure (e.g., brain imaging results confirming the presence of brain damage). The second vertical column includes those patients that are determined by the index measure to not have the condition in question. In research studies this is often classified as the control group. The first horizontal row includes those subjects that have been identified by the test under study as meeting the criteria for identification. For example, in cognitive screening tests this would include those subjects whose scores were below a specified cutoff, thereby indicating the presence of cognitive impairment. The second horizontal row includes those subjects whose scores on the test in question do not identify them as impaired. From this table a variety of ratios can be described, including the sensitivity (true positive rate) and the specificity (true negative rate) of the test in question (see Appendix D).

Sensitivity is described as the ability of the test to detect impairment when according to the index measure the individual is actually impaired (Haynes, 1981). The term "sensitivity is also synonymous with the true positive rate (Baldessarini, Finklestein, & Arana, 1983). Specificity is the ability of the test to correctly predict the absence of the impairment or disease as established by the lack of impairment or disease on the index measure. Specificity is synonymous with the true negative rate. Haynes (1981) describes the sensitivity and specificity as "stable measures" because these rates do not change when different proportions of patients and controls are tested (i.e., different base rates). The false positive rate and the false negative rate are also stable measures as they are simply the inverse of the specificity and the sensitivity ratios, respectively. It is very important to know the sensitivity and the specificity of a given measure, so that the clinician can judge the ability of the test to identify those who are truly impaired or ill, as well as those who are truly well. However, the clinician may not always have the results of an index measure to compare results. This is especially true in a situation when a cognitive screening measure might be employed.

In those situations, the clinician would be at least as interested in determining the likelihood that the disease or impairment is actually present given a positive or negative result on the test in question (Engelhart, Eisenstein, & Meininger, 1994). In other words, the clinician would be interested in the predictive or diagnostic utility of the test result (Baldessarini, Finklestein, & Arana, 1983). Positive Predictive Power (PPP) is the term used to describe the likelihood that the impairment or disease is actually present given a positive identification on the test (e.g., scored below cutoff for cognitive impairment). This is the ratio of true positive results to all the positive results obtained from the test in question. Negative Predictive Power (NPP) is the term used to describe the likelihood that the subject does not have the disease or impairment when the test results are negative. This is the ratio of true negative results to all the negative results obtained from the test in question (Baldessarini, Finklestein, & Arana, 1983; Engelhart, Eisenstein, & Meininger, 1994). The PPP and the NPP are described by Haynes (1981) as possessing "frequency-dependent properties" because these values vary depending upon the proportion of impaired to well individuals being tested. That is to say, these rates will change depending on the base rates of the ailment in question in any given population. Therefore, the PPP and NPP results obtained for a given test only apply to those clinical settings with similar base rates for the disease or impairment in question.

Haynes indicated that "the positive predictive value falls and the negative predictive value rises when a diagnostic test developed for patients with high prevalence of the target disorder is subsequently applied to patients with a lower prevalence of the disorder" (1981, p. 706). It is not uncommon to see this difference in prevalence rates between the research conducted on a test as compared to its use in general clinical practice. In research situations, the test subjects are often matched one for one with a control group, resulting in a 50 percent prevalence rate. However, very few clinical settings have a prevalence rate of 50 percent for the condition to be identified by a test. In an hypothetical example provided by Baldessarini, Finklestein, and Arana, (1983) when the prevalence rate was 50 percent the PPP was 93 percent, while the NPP was 76 percent (sensitivity and specificity of the test was fixed at 70 % and 95 %, respectively). By the time the prevalence rate had dropped to only 1 percent of the population the PPP had dropped to 12 percent and the NPP had risen to 99.7 percent. PPP and NPP are also affected when the prevalence is held constant, but the sensitivity or specificity is modified. Under this situation Baldessarini, Finklestein, and Arana conclude that:

> the NPP is highest when sensitivity is high (i.e., when the rate of false negatives is low), and the PPP is high only when the specificity is high (i.e., the rate of false positives is low). Although the PPP is high when a test is highly specific, this power is markedly decreased as prevalence decreases. These considerations are consistent with the impression that highly specific tests (those having low false-positive rates), with even moderate sensitivity, are especially useful when results are positive and when the prevalence of the index condition is high (as in a specialty practice). Conversely, highly sensitive tests (those having low false-negative rates), with at least moderate specificity, are powerful when results are negative and when the prevalence of an illness is low (as in a screening program). Such a highly sensitive test can be useful in excluding a diagnosis.(1983, p.572)

This information helps to emphasize the importance of understanding the origin and development of the reported predictive power values, and the resulting implications for clinical settings that vary from the reported research.

Functional Assessment of Skills and Abilities

There is increasing pressure to establish the efficacy of rehabilitation programs. Administrators are requesting that information regarding a program's cost effectiveness and ability to achieve the stated rehabilitation goals be delineated and quantified. One method that can be utilized in an attempt to meet these demands is through the assessment of an individual's functional skills and abilities. Batavia (1988; as cited in Wilkerson, Batavia, & DeJong, 1992), in his review of payment for rehabilitation services, indicated that "payment per level of functional status" was the most theoretically attractive because it was most consistent with the rehabilitation goal of enhancing the individual's functional status. If this type of rehabilitation payment system is to be adopted, it is critical that reliable and valid measures of functional status be developed.

Functional assessments (Crewe & Dijkers, 1995) attempt to directly measure, in a natural setting, a person's ability to carry out meaningful behavioral tasks or to observe his or her actual performance of these tasks. In comparison, psychometric norm-referenced testing often utilize questions and tasks that allow inferences and predictions to be made about how a person will perform in different situations. Further, functional assessments typically allow for repeated assessments with the same instrument as a means of tracking the individual's change over time. A variety of instruments have been developed over the years in an attempt to produce meaningful measures of functional skills and abilities. Although many different functional status measures have been developed, one measure that has become commonly used for program evaluation and rehabilitation outcome studies is the Functional Independence Measure (FIM) (Crewe & Dijkers, 1995). This scale is typically used in rehabilitation in-patient programs to assess the extent of the patients' physical disabilities and the ability to perform self-care activities.

The FIM was designed as a basic indicator of severity of disability in order to track a patient's progress from entry into a hospital setting through to discharge and follow-up, thereby providing data on the effectiveness and efficiency of rehabilitation. In 1984, a task force was established at the Department of Rehabilitation Medicine, School of Medicine, State University of New York at Buffalo to develop a uniform system to document the severity of an individual's disability (Guide for the Uniform Data Set for Medical Rehabilitation [UDSmr], 1993). The development of the FIM was intended to produce a rating scale instrument that could be quickly and uniformly administered, and that "would include only key patient functional attributes - those that were common and useful, that would be discipline free, and acceptable to clinicians, administrators, and researchers" (Guide for the Uniform Data Set for Medical Rehabilitation [UDSmr], 1993, p.I-1). Items and concepts from numerous other functional assessment measures, such as the Katz Index of ADL, PULSES, Kenny Self-Care Evaluation, and the Barthel Index, were included in the development of the FIM. (Hall, Hamilton, Gordon, & Zasler, 1993). The Functional Assessment Measure (FAM) was developed as an adjunct to the FIM for use with brain-injured individuals, and is reported to add sensitivity to the functional assessment of individuals in the post-acute rehabilitation stage (Hall & Johnston, 1994). The FAM was developed as an expanded version of the Functional Independence Measure (FIM), which was developed as part of the Uniform Data System for Medical Rehabilitation (UDS).

The FIM + FAM consists of a total of 30 items (see Appendix E), which are rated on a 7 point scale (see Appendix F). The 30 items in the FIM+FAM consist of the original 18 items from the FIM, along with an additional 12 items that focus primarily upon the cognitive and psychosocial elements of an individual's disability. Performance on any of the 30 items is designed to assess whether an individual can accomplish a given activity independently or if the individual requires assistance from another person and the amount of time required of the helper to accomplish that task. The items are to be rated for what the person actually does, as opposed to what the evaluators believe the person is capable of or should be doing (UDSmr, 1993).

The FIM has been utilized by a number of researchers, as a means of assessing functional abilities and level of care needed during and after rehabilitation with stroke patients. One study assessed the ability of the FIM to accurately predict the burden of care in discharged stroke rehabilitation patients (Granger, Cotter, Hamilton, & Fielder, 1993). Burden of care was determined by the number of minutes of care per day required from a caregiver. For the 21 subjects in this study it was reported that a one point change in total FIM score was equivalent to an average of 2.19 minutes of care per day required from another individual. The researchers reported that the FIM produced a finer gradation of minutes of care than the other measures used in their study. The FIM was also reported to help predict the patient's general life satisfaction. Granger, et al. (1993) concluded that, although the FIM did overlap with some information gathered from other sources, it did provide unique and important predictive information in regard to burden of care and general life satisfaction.

Alexander (1994) studied the relationship between the side of stroke, age, and admission functional levels against discharge setting and change in functional ability for 520 stroke patients. It was found that all patients younger than 55 years old were discharged home regardless of initial severity of disability. Also, most patients admitted with moderate levels of functional disability were discharged home regardless of age. The FIM scores were found to be useful when attempting to predict patient disposition. For example, it was reported that low admission FIM scores and poor FIM-change during rehabilitation commonly resulted in a nursing home discharge (Alexander, 1994). It was concluded that "standard clinical measures available at rehabilitation admission carry enough predictive power to define management strategies for stroke survivors" (Alexander, 1994, p. 128).

The FIM was analyzed to help determine those variables that best predicted discharge home for 282 stroke patients (Wilson, Houle, & Keith, 1991). Seventy-five percent of their study population returned home. A regression model was developed to predict discharge disposition. This model used FIM admission and discharge total scores, living arrangement before the stroke, and length of stay. The use of the regression model correctly predicted discharge outcome for 83 percent of the subjects, which is a modest increase from the base rate of 75 percent. Patients who returned home had higher total FIM scores at admission and discharge, and they remained in rehabilitation longer. Those patients who returned home had average FIM scores in the moderate to minimal assist range by the time of discharge. It was reported that other variables, such as age, side of hemiplegia, sex, and onset-admission interval did not make significant contributions toward predicting discharge home.

In a study of 7,905 stroke patients drawn from the Uniform Data System, the relationship between subject variables and program variables was examined (Granger, Hamilton, & Fielder, 1992). Subject variables included age and side of body affected, while the program variables included length of stay, levels of functional ability, and rates of community discharge. In general, bilaterally impaired subjects were reported to have the lowest FIM scores for admission and discharge, as well as the least change during rehabilitation across all age groups. Lengths of stay and discharge outcome were similar for left and right hemiparetics. Increasing age adversely affects recovery after stroke, shortens the stay in rehabilitation, and lowers the probability of discharge into the community, according to the results of this study. The results of the above described studies indicate that the FIM can provide useful predictive information about the rehabilitation needs and discharge outcome of stroke patients.

Individuals who have suffered traumatic brain injury (TBI) are another patient population that has been studied using the FIM. In a study of 109 multi-trauma patients in an acute care hospital setting, it was discovered that patients who were discharged home exhibited significantly greater FIM score improvements per day compared to those patients who were discharged to a rehabilitation setting (Emhoff, McCarthy, Cushman, Garb, & Valenziano, 1991). There were no significant differences in regard to age or gender between the groups for those who went home and those who were discharged to rehabilitation. However, it was found across groups that functional level as measured by the FIM, determined discharge disposition. Those patients who were discharged home also exhibited significantly greater overall gains on the FIM from admission to discharge. The researchers concluded that the FIM "produced a very good measure of the patient's total function, tracked progress (or lack of it) through the acute hospitalization, and correctly categorized and quantitated dysfunction (cognitive and physical) as discharge planning was being done" (Emhoff, McCarthy, Cushman, Garb, & Valenziano, 1991, p. 1230).

Three hundred and twenty eight TBI patients with FIM scores of 18 at admission were studied in an analysis of patients admitted to acute care rehabilitation hospitals and clinics (Whitlock & Hamilton, 1995). Discharge FIM scores averaged 53, although there appeared to be a bi-modal distribution with approximately one quarter of the subjects showing no change in FIM scores from admission to discharge. More than two thirds of the subjects that exhibited FIM score improvements were discharged home. This study reported that a simple relationship between time and FIM scores was not apparent (i.e., there was no trend toward improvement with time in the acute care setting). However, the researchers predicted that administering the FIM at regular intervals throughout rehabilitation would provide a useful means of tracking patterns of recovery and assessing such factors as pathology, time, environment, and therapeutic interventions (Whitlock & Hamilton, 1995).

A study of 91 TBI patients in a rehabilitation program analyzed a number of variables with regard to the relationship between the FIM and other pre-admission and outcome variables (Cowen et al., 1995). Patients in the severely impaired group (Glasgow Coma Score = 3 to 7) exhibited significantly lower scores at admission and discharge on both the cognitive and motor FIM subscales. It was reported that the most powerful predictor of length of stay (LOS) in rehabilitation and the changes over the course of rehabilitation was the FIM motor score. The researchers reported that longer acute care LOS was associated with significantly lower FIM scores on admission to rehabilitation, even when other variables such as Glasgow Coma Scores, CT findings, etiology, and age were controlled. On the other hand, longer LOS in rehabilitation was related to greater gains in the FIM motor and cognitive subscales. These results provide further evidence for the utility of the FIM in rehabilitation settings for helping predict rehabilitation outcomes and charges.

The FIM has also proved useful in attempting to predict long-term outcome after discharge from rehabilitation. In a study of 116 TBI survivors, who were given a follow-up assessment two years after injury and rehabilitation, quality of life factors were rated (Webb, Wrigley, Yoels, & Fine, 1995). The researchers used a path analysis to infer direct and indirect effects of the independent variables (e.g., employment, functional ability, and family support) on quality of life. Functional independence (as measured by the FIM) was reported to have a significant direct relationship to employment status. In turn, employment status was reported to have the largest direct relationship to quality of life ratings. It was also reported that functional independence at 24 months (as measured by the FIM) had a significant direct relationship to quality of life ratings. Participation in rehabilitation had a relationship to functional independence at 24 months, and therefore, presumably had an indirect relationship to quality of life. These results suggest that the FIM could contribute useful information about a patient's post-rehabilitation employment status and quality of life.

The studies presented here on the stroke and TBI patients serve to emphasize the different ways in which the information derived from the FIM can assist those agencies and professionals involved in a patient's care. It has proven useful from initial admission to acute care settings through to long-term follow-up after rehabilitation discharge. It is important to demonstrate that the psychometric characteristics of this measure are sound considering the extent to which it is being used at the present time in a variety of settings.

FIM+FAM Test Reliability and Validity

A necessary and critical step in the development of any test instrument is to establish its reliability and validity. As defined by Anastasi (1988, p. 110), "all types of reliability are concerned with the degree of consistency or agreement between two independently derived sets of scores". Many types of reliability are described, including test-retest, split-half, and alternate-form reliability. To establish test validity in the context of behavioral testing, Anastasi states that "all procedures for determining test validity are concerned with the relationships between performance on the test and other independently observable facts about the behavior characteristics under consideration" (1988, p. 139). The three general types of test validity commonly identified are content-related, criterion-related, and construct-related validity. However, before proceeding to testing the validity of a measure it is important to establish its reliability. The reason for this is that the reliability of a measure sets a limit on the validity of that measure. Although a measure might be highly reliable across time or within itself, it may still not measure what it was designed to assess for a number of reasons (e.g., poorly selected items that due not effectively assess the construct in question). On the other hand, if a measure is not reliable then its validity will be constrained by the low internal consistency and/or weak stability over time.

One of the most comprehensive studies, in terms of population size, on the FIM's inter-rater reliability was conducted on 263 patients undergoing medical rehabilitation at 21 U.S. hospitals (Hamilton, Laughlin, Granger, & Kayton, 1991). Two or more pairs of clinicians assessed each of the patients. An intraclass correlation coefficient (ICC) ANOVA was calculated for each of the FIM subscales (i.e., Self-Care, Sphincter Control, Transfers, Locomotion, Communication, & Social Cognition), and for the total FIM score. The ICC's ranged from 0.97 (Total FIM score) to 0.93 (locomotion). The inter-rater reliability for each of the items within the FIM was calculated using unweighted Kappa. These values were reported to range from 0.76 to 0.61 (mean of 0.71). Based upon these values the authors concluded that the FIM possesses good clinical inter-rater agreement.

In a study of 38 stroke survivors during a follow-up assessment, the ICC for the total FIM was reported to be 0.96, as rated by two trained raters (Segal & Schall, 1994). Segal and Schall (1994) also had the patient and their caregiver provide ratings on the FIM. It was found that the overall ICC between the caregiver and the patient was 0.87, with a considerably better overall ICC for the physical items (i.e., 0.91) as opposed to a lower overall ICC (i.e., 0.60) on the scales assessing cognitive dimensions.

In a study of forty stroke patients, the follow-up inter-rater reliability between phone interviews and in-person assessments was measured (Smith, Illig, Fielder, Hamilton, & Ottenbacher, 1996). Two separate raters conducted either a telephone FIM assessment or an in-person assessment for each subject in an alternating pattern across subjects. The total FIM ICC was 0.97, with the FIM subscale ICC's ranging from 0.85 (Communication) to 0.98 (Self-Care). The Social Cognition subscale was not included within this part of the analysis because of a lack of variability in the obtained scores resulting in an artificially low ICC value. The authors concluded that good intermodal agreement was found between the telephone and in-person follow-up assessments for patients with effective communication skills.

The inter-rater reliability studies performed on the FIM have generally produced higher inter-rater reliability values for the items and subscales that focus upon motor skills, with lower values for the psychosocial items and subscales. In order to address the psychometric structure of the 18 FIM items, Linacre et al. (1994) conducted a Rasch analysis on the results of 14,799 patient evaluations. A Rasch analysis helps determine if a psychometric scale can be transformed to fit three assumptions required for interval scales. These three concepts are: additivity, order, and unidimensionality (Andiel, 1995). Additivity refers to the concept that there are equal distances between the units of measurement for the scale being used. Order refers to the concept that a higher number indicates a greater value of whatever is being measured. Finally, unidimensionality refers to the idea that the instrument must be measuring only one construct or feature.

Initially, Reid and Bonwich (1989), in a factor analytic study of the FIM found that it was not unidimensional. Subsequently, the Rasch analysis of all 18 items of the FIM, conducted by Linacre et al. (1994, p. 132), supported this finding by indicating that "the FIM items detect two substantially different aspects of disability: motor function and cognitive function". This distinction between the motor and cognitive items was further supported in an analysis of FIM results from 27,699 subjects with various impairments (Heinemann, Linacre, Wright, Hamilton, & Granger, 1993). They found that the motor items were generally better predictors of length of stay than the cognitive items across all impairment groups, although the cognitive items provided unique predictive contributions for some groups (e.g., traumatic brain dysfunction).

At this time there is very little published research on the FIM+FAM. Hall, Hamilton, Gordon, and Zasler (1993) report some preliminary work on the inter-rater reliability of the FIM+FAM. They utilized narratives based upon physical and history reports from three case studies. These three case report narratives were then rated by a total of 20 raters providing ratings for admission, discharge, and follow-up. Overall FIM agreement (i.e., "the number of identical ratings across every item") was reported to be high (i.e., 88%), although FIM+FAM agreement was less satisfactory at 67 percent. When the ratings at admission were analyzed, the percentages were even lower for both the FIM (81%) and the FAM (55%). One possible explanation for this discrepancy, put forward by the researchers, was that none of the raters had been trained in scoring the FIM+FAM items. Of further concern, and probably more important, is the fact that the method of determining the "percent of agreement" was not stipulated in their report.

Wass (1996b) conducted an inter-rater reliability study on the FIM+FAM within a post-acute care brain injury rehabilitation program. A total of 53 patients (41 males, 12 females) with closed head injuries were scored by varying numbers of treatment team members over a 12 week period. Bartko and Carpenter's (1976) intraclass correlation coefficient via the analysis of variance was utilized. As identified by Cicchetti (1991), this model of the ANOVA ICC permits for determining reliability when the numbers and specific sets of raters may vary at each assessment.

The following table (Table 2.) presents the inter-rater reliability of the FIM+FAM items, ranked from the most to least reliable.

Insert Table 2. here

According to the criteria established by Cicchetti and Sparrow (1981), all but one of these items were within the Good to Excellent range for determining the clinical significance of an ICC value. The FIM item, Social Interaction, fell within the Poor range, with an ICC value of 0.356.

The overall mean ICC for all items was .8288 (SD = .1284). The mean ICC for the FIM items alone was .8538 (SD = .1464), and for the additional FAM items was .7912 (SD = .0255). The mean values were also computed for the FIM+FAM scale items from the first three domains, which focus primarily upon motor/physical care issues (items 1 to 16), and for the last three domains, which focus upon psychosocial/cognitive issues (items 17 to 30). The mean for the motor/physical care items was .9095 (SD = .0597), and for the psychosocial/cognitive items was .7365 (SD = .1244). Wass (1996b) summarizes the results of this study by stating that items in the motor/physical care domain were rated by fewer people and had inter-rater reliability values within the excellent range. Items within the final three domains were typically rated by more people and had inter-rater reliability values within the good range. It was concluded that these results provided further evidence to support the clinical utility of the FIM+FAM.

Hall, Hamilton, Gordon, and Zasler (1993) analyzed the relationships among three functional assessment indices (the Disability Rating Scale [DRS], the FIM, and the FIM+FAM), and six indices of severity of injury (i.e., Glascow Coma Scale [GCS], length of coma, length of post-traumatic amnesia [PTA], Revised Trauma Score, computed tomography [CT] pathology, and the Rancho Los Amigos Levels of Cognitive Functioning Scale [LCFS]). Although the sample sizes were reported to be variable because of missing data for some analyses, there were a total of 332 subjects pulled from a national database. Inter-correlations among the functional assessment indices were highly significant for all three of the disability measures, and a two-factor principal components analysis accounted for 92 percent of the variance (a cognitive factor and a motor factor). A principal components analysis of the severity of injury measures found that the trauma score and GCS were distinctly separate from the LCFS, PTA, and length of coma factor. The CT pathology measure did not load on any of the factors and was found to have the weakest inter-correlation with the other severity indices.

All three functional assessment indices were significantly associated with most of the severity of injury measures, with all three indices most strongly associated with the LCFS, and none being significantly associated with CT pathology. In regard to the FIM and FIM+FAM, the association with the LCFS was greatest for the FIM+FAM cognitive measures. The researchers concluded that this study provided evidence for the concurrent validity of these measures because of the strong correlations between the three functional assessment indices studied (Hall, Hamilton, Gordon, & Zasler, 1993). They further state that the FIM+FAM appears to exhibit increased sensitivity over the other measures for post-acute rehabilitation functional assessments, although the time for completion (e.g., approximately 35 minutes) makes it somewhat more unwieldy. The researchers argue that the FIM+FAM is reported to be well suited as a measure of the finer details of functional status and outcome. They conclude that it is important to determine the reliability of the FAM, in large part because of the increasing number of requests being made from rehabilitation settings to use the FAM (Hall, Hamilton, Gordon, & Zasler, 1993, p. 72).

It is important to recognize that rating a patient's cognitive functioning according to *functional ability*, as determined by the FIM+FAM, is likely to present difficulties. For example, a neuropsychological assessment will attempt to evaluate a patient's cognitive strengths and weaknesses, and will provide information regarding their levels of cognitive *impairment*, if present. However, their actual *functional* presentation on a day-to-day basis in the rehabilitation setting can be a result of a large number of factors (e.g., pre-morbid personality structure, level of social support, amount of structure provided in the rehabilitation setting) that may be external to their level of cognitive impairment. These factors will help determine the degree to which *disabilities* and *handicaps* may be manifest. It is not expressly stated which level of functioning the FIM+FAM is attempting to delineate. The scoring criteria suggest that it is more focused upon the disabilities and handicaps level.

These issues could help to partially explain the lower inter-rater reliability scores found in the Psychosocial-Cognition items of the FIM+FAM. Some professionals may be interpreting and scoring a patient based upon the test results specific to their profession (e.g., Speech and Language, Neuropsychology) and day-to-day therapeutic interactions that focus upon remediating those impairments. Other professional groups would not have first hand knowledge of these cognitive impairment test results and would be determining the patient's FIM+FAM scores on their personal interactions. These fundamental differences in perspective about how the patient presents are likely to affect the agreement between raters on a scale such as the FIM+FAM, and these differences would be most apparent on the cognitive items.

Ecological Validity of the FIM and the FIM+FAM

The studies conducted to date on the FIM, as well as the FIM+FAM, address a number of concerns expressed by Franzen and Wilhelm (1996) about the psychometric properties of functional ability instruments (see the discussion above on ecological validity). The first concern was the subjective nature for scoring most functional ability measures. Franzen and Wilhelm (1996) suggest that strong inter-rater reliability would help provide evidence of sufficiently specific scoring instructions, thereby allowing for consistent ratings. The studies described above on the inter-rater reliability of the FIM and the FIM+FAM provide consistent evidence to indicate that different evaluators can reliably score the scoring system. A second concern, which is related to the first issue, is the complexity of the behaviors being assessed on any one item. Rating activities that encompass a large number of different behaviors are more likely to be scored substantially different by different evaluators, since there is a broader range of behaviors that could be the focus of analysis. Once again, Franzen and Wilhelm (1996) suggest that evidence of high inter-rater reliability could address this concern. The excellent range inter-rater reliability scores obtained for the motor/physical care items, as compared to the generally good range psychosocial/cognitive item scores, suggest that there could be problems in how the psychosocial/cognitive items are described and rated in the scoring criteria. However, almost all of the psychosocial/cognitive items are still within a very acceptable inter-rater reliability range, thereby supporting their inclusion in the overall measure.

The third concern expressed by Franzen and Wilhelm (1996) was a lack of information about the scale properties of most of the functional assessment instruments. This concern has been well researched for the FIM and some literature is beginning to provide evidence for the FIM+FAM. There is evidence available from Rasch analyses to support the use of the FIM and the FIM+FAM as an ordinal scale. Factor analytic studies suggest that the FIM is bi-dimensional, with the motor items making up one factor, and the psychosocial items composing the second factor.

The final concerns involve the relationship between the items on the functional assessment instrument and the actual behaviors as expressed outside of the assessment situation. One way in which this issue is addressed in the FIM and the FIM+FAM is by having the professionals involved in the individual's care rate a person based upon their interactions with them in day-to-day activities and treatment. This is in contrast to having each item specifically tested in a controlled setting. The research studies also provide further evidence to support the relationship between these measures and actual behaviors. Studies that relate the person's performance on the FIM and the FIM+FAM to such factors as employment, level of care after rehabilitation, and general quality of life indicators provide evidence to indicate that these instruments can explain at least a portion of the variance in these 'actual behaviors'. In conclusion, it is reasonable to state that the veridicality of the FIM and the FIM+FAM has been established based upon the body of literature available.

Purpose

The BNIS is a recently developed cognitive screening instrument that has not been subjected to extensive testing of its psychometric properties in clinical settings independent of its site of origin. This study will provide further information on the construct-related and ecological validity of the BNIS, from a patient population that was independent of the data gathered for the published normative studies. The sample of patients for this study was obtained through archival records at the Brain Injury Rehabilitation Program (BIRP) in Alberta Hospital Ponoka. The construct-related validity of the BNIS will be analyzed through its relationship with cognitive measures that were administered as part of a full neuropsychological evaluation. These analyses will be an extension of the validity studies reported by Prigatano, Amin, and Rosenstein (1993b).

The ecological validity of the BNIS will be examined through its relationship with the Functional Independence Measure (Guide for the Uniform Data Set for Medical Rehabilitation, 1993) and its adjunct the Functional Assessment Measure (FIM+FAM) (Hall & Johnston, 1994). As described in the literature review, the veridicality of the FIM+FAM has been demonstrated through its relationship with various indicators of daily living skills and performance. The literature review on the FIM+FAM research studies also addresses the concerns expressed by Franzen and Wilhelm (1996) about the psychometric properties of functional ability measures. Therefore, the FIM+FAM will be used as a measure of functional day-to-day performance. The ecological validity of the BNIS will be analyzed through its relationship with the FIM+FAM, based on data gathered from the BIRP patient population.

Hypotheses

Hypothesis one.

Hypothesis one states that there will be a significant positive correlation between the BNIS Total and Subscale scores (i.e., Speech and Language Functions, Orientation, Attention/Concentration, Visuospatial and Visual Problem Solving, Memory, Affect, and Awareness versus Performance) and selected neuropsychological test items. The neuropsychological items included the WMS Logical Memory I & II, and Visual Reproduction I & II subscales; the WAIS-R Vocabulary, Information, Similarities, Picture Completion, Block Design, Object Assembly, and Digit Symbol subtests; Trail Making Test A & B; Grooved Pegboard Test; Halstead Category Test (HCT); Wisconsin Card Sorting Test (WCST) categories and perseverative responses; and the Controlled Oral Word Association (FAS) Test (COWAT). Most of the neuropsychological measures included in this analysis were the same measures included in the correlational study reported by Prigatano, Amin, and Rosenstein (1993b). The additional measures included in this analysis were the HCT (Halstead, 1947; Reitan & Wolfson, 1985), the WCST (Berg, 1948; Heaton, 1981), the COWAT, the Grooved Pegboard Test, and the Vocabulary and Object Assembly subtests from the WAIS-R. The HCT was included because of its generally recognized sensitivity to the presence of brain damage regardless of location (Lezak, 1995). The WCST was included because of its recognized sensitivity to the presence of executive function disorders. The COWAT was included as a verbal measure sensitive to executive dysfunction. The Grooved Pegboard Test was included as a measure of right versus left hand performance. It would be expected that right hand performance would correlate strongest with the BNIS Speech & Language and the Memory Subscales.

Performance for the left hand would be expected to correlate strongest with the Visual Spatial & Visual Problem Solving and the Affect Subscales. These relationships will provide support for the construct-related validity of the BNIS.

Hypothesis two.

Hypothesis two states that the BNIS Speech and Language Subscale items will predict performance on the WAIS-R Verbal IQ measure. The WAIS-R Verbal IQ was selected as a global measure of language-related ability. In order to maintain a minimum of at least five subjects per independent variable (Tabachnick & Fidell, 1989, p. 129) some of the BNIS Speech and Language items will be combined to form composite measures. The first composite measure is composed of the Fluent, Paraphasic, and Dysarthric items, and will be called the BNIS Expression composite measure. These three items were combined because they assess aspects of the patient's expressive language ability as determined by the examiner throughout the course of the testing. The second composite measure is composed of the Writing Copy and Writing to Dictation items, and will be called the BNIS Writing composite measure. The third composite measure is composed of the Spelling – Irregular and the Spelling – Phonetic items, and will be called the BNIS Spelling composite measure. Finally, the Arithmetic – Alexia and the Arithmetic – Dyscalculia items will compose the Arithmetic composite measure. This combination of scores will provide a ratio of 8.25 subjects per independent variable. These relationships will provide support for the construct-related validity of the BNIS.

Hypothesis three.

Hypothesis three states that the BNIS Visual Spatial and Visual Problem Solving Subscale items will predict performance on the WAIS-R Performance IQ measure. These relationships will also provide support for the construct-related validity of the BNIS.

Hypothesis four.

Hypothesis four states that the patient's discharge BNIS Total Score and the Subscale scores (i.e., Speech and Language, Orientation, Attention/Concentration, Visuospatial & Visual Problem Solving, Memory, Affect, and Awareness versus Performance) will exhibit a significant positive correlation with the discharge Total FIM+ FAM score. Further, the FIM+FAM Psychosocial/Cognitive subscale total score (i.e., items 17 to 30) will also exhibit a significant positive correlation with the BNIS Total Score and Subscale scores. However, the FIM+FAM Motor subscale total score will exhibit a weaker positive correlation with the BNIS Subscale scores. These relationships will provide support for the ecological validity of the BNIS.

Hypothesis five.

Hypothesis five states that the discharge BNIS Speech and Language Subscale items will predict performance on the discharge FIM+FAM Communications subtotal. The five FIM+FAM Communications items (i.e., Comprehension, Expression, Reading, Writing, and Speech Intelligibility) will be summed and used as the dependent variable in this analysis. In order to maintain a minimum of at least five subjects per independent variable (Tabachnick & Fidell, 1989), the same summing procedure as described in Hypothesis Two for the BNIS Speech and Language items will be applied. This combination of scores will provide a ratio of 10.6 subjects per independent variable. The predictive relationship between the BNIS items and the FIM+FAM Communication subscale will provide support for the ecological validity of the BNIS.

Hypothesis six.

Hypothesis six states that the three BNIS Orientation Subscale discharge items will predict scores on the discharge FIM+FAM Orientation measure. This relationship will provide support for the ecological validity of the BNIS.

Hypothesis seven.

Hypothesis seven states that the three BNIS Attention/Concentration Subscale discharge items will predict scores on the discharge FIM+FAM Attention measure. This relationship will provide support for the ecological validity of the BNIS.

Hypothesis eight.

Hypothesis eight states that the BNIS Visual Scanning, Visual Sequencing, Pattern Copying, Pattern Recognition, and Awareness versus Performance discharge items will predict performance on the discharge FIM+FAM Problem Solving measure. The Visual Object Recognition item from the BNIS will not be included because the task demands (i.e., a two-choice pointing task to identify a pencil) are more consistent with auditory comprehension and visual perception. The BNIS Awareness versus Performance item will be included because of its relation to higher-order cognitive processes that can impact on executive functions. The relationship between these BNIS items and the FIM+FAM Problem Solving measure will provide support for the ecological validity of the BNIS.

Hypothesis nine.

Hypothesis nine states that the discharge BNIS Memory items will predict performance on the discharge FIM+FAM Memory measure. The BNIS Memory items include the Learning and Memory item and each of the three words in the delayed recalled task. This relationship will provide support for the ecological validity of the BNIS.

Hypothesis ten.

Hypothesis ten states that the discharge BNIS Affect items will predict performance on the discharge FIM+FAM Emotional Status measure. The BNIS Affect items included Affect Expression, Perception of Facial Affect, Affect Control, and Spontaneous Affect. This relationship will provide support for the ecological validity of the BNIS.

Hypothesis eleven.

Hypothesis eleven states that the discharge BNIS Subscale totals (i.e., Speech and Language, Orientation, Attention/Concentration, Visual Spatial/Visual Problem Solving, Memory, Affect, and Awareness versus Performance) will predict performance on the FIM+FAM Psychosocial-Cognition items subtotal. This relationship will provide support for the ecological validity of the BNIS.

CHAPTER II

METHODOLOGY

Participants

The data gathered for the purposes of this study is from an archival database gathered from patients who had been admitted to the Alberta Hospital Ponoka - Brain Injury Rehabilitation Program (BIRP) in Central Alberta. The BIRP focuses upon rehabilitation for post-acute care brain injured adults. The data had been accumulated over the past two years, from 1994 to 1996 in the course of day to day clinical intervention. The medical staff described the majority of the patients as having sustained moderate to severe brain injuries. All patients included in this study were in-patients.

In the first part of the study, analyzing the relationship between the BNIS and neuropsychological assessment results, there are seventy-two patients. Only those patients who were discharged, and had received both a BNIS assessment and a full neuropsychological assessment were included in this part of the study. This group had a mean age at testing of 36.5 years (S.D. 13.0; range = 17 to 69), and was composed of 49 males and 23 females. The mean education level was 11.7 years (S.D. 2.1; range = 6 to 18 years). Sixteen percent of the patients in this part of the study were tested within three months of injury, forty-eight percent were between three and twelve months post injury, and thirty-six percent were past twelve months.

There are eighty-five patients in the second part of the study analyzing the relationship between the BNIS and the FIM+FAM discharge results. This group had a mean age at testing of

39.3 years (S.D. 14.4; range = 17 to 70), and was composed of 55 males and 30 females. The mean education level was 11.0 years (S.D. 2.5; mode = 12 years (23.5 %); range = 5 to 18 years).

Materials

The FIM+FAM (see Appendix E and F) was administered to all patients included in this study. As part of the regular tracking procedure in the Neuropsychology Program, each patient had been assessed with the BNIS (see Appendix A and B). The patients included within the first part of this study had been assessed with a comprehensive neuropsychological test battery. This battery included the Wechsler Adult Intelligence Scale - Revised (Wechsler, 1981); subtests from the Wechsler Memory Scale - Revised (Wechsler, 1987); the COWA Test (FAS); the WCST (Heaton, 1981); and selected measures from the Halstead-Reitan neuropsychological test battery including the Halstead Category Test and the Trail Making Test, Part A and B (Reitan & Wolfson, 1985); and the Grooved Pegboard Test.

Procedures

Within the first one to two weeks of admission to the BIRP each patient was rated on the FIM+FAM by the treatment team members (see Appendix E and F). A second rating on this scale was conducted approximately 45 days after the initial FIM+FAM. Subsequently, every 3 months the patient would be re-evaluated by the treatment team on the FIM+FAM until discharge from the BIRP. If the patient was capable of receiving the full screening examination, they would typically be tested within one week of the treatment team rating them on the FIM+FAM. A primary component of the screening examination included the BNIS. The patient's ability to comply with the screening examination, along with their quantitative performance, was a significant factor in determining when the patient was capable of receiving a full neuropsychological evaluation. Those patient's who were deemed appropriate, as determined by the supervising clinical neuropsychologist, were administered a full neuropsychological battery including the measures described above. The neuropsychological evaluation coincided with the first administration of the BNIS for fifty-six of the 72 patients, with the remaining (i.e., 16 patients) neuropsychological evaluations coinciding with the second administration of the BNIS.

For the second part of this study, the final discharge BNIS assessment and FIM+FAM results were utilized. The FIM+FAM discharge assessment reflected the treatment team's evaluation of the patient's functional performance at the conclusion of active rehabilitation, and just prior to their re-entry into the community. Community placements upon discharge ranged from group homes and various managed-care facilities to independent living. These two measures were completed within approximately two to three weeks of each other surrounding the discharge of a patient from the BIRP. Sixty-two of the 72 subjects from the first part of the study (comparing the BNIS to the neuropsychology data) were included in the second part of the study.

CHAPTER III

RESULTS

The data were entered into a microcomputer database and were analyzed using the Statistical Package for Social Sciences for Windows, Release 7.5 (Norusis, 1997). The first series of analyses describes the relationship between the BNIS and neuropsychological measures. This is presented as evidence of the construct-related validity of the BNIS. The second series of analyses describes the relationship between the BNIS and the FIM+FAM. This is presented as evidence of the ecological validity of the BNIS. Results are presented in terms of tests of each hypothesis.

Construct-related validity of the BNIS

Hypothesis One.

The first analysis conducted in this study examined the relationship between neuropsychological measures and the BNIS Subscales and BNIS Total Score. Pearson correlation coefficients between these variables are provided in Table 3. The BNIS Total score was correlated with all neuropsychological measures, with most correlations being significant at the 0.001 level.

Table 3

Pearson Correlation Coefficients Between the BNIS and Neuropsychological Measures

	BNIS Total	Speech & Language	Orientation	Attention / Concentration	Visual Spatial & Problem Solving	Memory	Affect	Awareness
Logical Memory I (n=62)	.39 **	.39 ***	.38 **	.31 *	.00 ns	.38 **	.16 ns	-.17 ns
Logical Memory II (n=62)	.42 ***	.38 **	.33 **	.24 ns	.12 ns	.40 ***	.19 ns	-.06 ns
Visual Reproduction I (n=67)	.41 ***	.31 **	.35 **	.16 ns	.36 **	.27 *	.38 ***	.07 ns
Visual Reproduction II (n=67)	.53 ***	.31 **	.41 ***	.41 ***	.38 **	.55 ***	.32 **	.09 ns
Block Design (n=69)	.52 ***	.34 **	.31 **	.30 *	.44 ***	.40 ***	.40 ***	.10 ns
Picture Completion (n=68)	.42 ***	.22 ns	.33 **	.21 ns	.37 **	.27 *	.46 ***	-.00 ns
Object Assembly (n=69)	.44 ***	.31 **	.31 *	.29 *	.38 ***	.26 *	.43 ***	-.03 ns
Digit Symbol (n=63)	.57 ***	.42 ***	.23 ns	.27 *	.53 ***	.44 ***	.46 ***	.09 ns
Information (n=68)	.52 ***	.49 ***	.45 ***	.34 **	.19 ns	.29 *	.49 ***	-.07 ns
Similarities (n=68)	.46 ***	.32 **	.35 **	.28 *	.30 *	.32 **	.42 ***	-.09 ns
Vocabulary (n=67)	.48 ***	.48 ***	.41 ***	.28 *	.24 *	.21 ns	.44 ***	-.15 ns
Trails A (n=72)	.48 ***	.42 ***	.20 ns	.27 *	.36 **	.23 *	.58 ***	-.00 ns
Trails B (n=64)	.55 ***	.42 ***	.26 *	.36 **	.42 ***	.36 **	.44 ***	.06 ns
Grooved Pegs Right (n=61)	.44 ***	.36 ***	.24 ns	.18 ns	.32 *	.33 **	.30 *	.09 ns
Grooved Pegs Left (n=61)	.35 **	.17 ns	.21 ns	.25 *	.33 **	.30 *	.38 **	.23 ns
Category Test (n=66)	.30 ***	.10 ns	.27 *	.24 *	.24 *	.27 *	.36 **	.12 ns
WCST Categories (n=71)	.45 ***	.26 *	.29 *	.25 *	.31 **	.39 ***	.49 ***	.13 ns
WCST Perseverative Responses (n=71)	.43 ***	.28 *	.20 ns	.22 ns	.31 **	.32 **	.57 ***	.09 ns
COWAT (FAS) (n=62)	.48 ***	.56 ***	.20 ns	.18 ns	.27 *	.24 ns	.35 **	.05 ns

ns = Not significant; * = $p < 0.05$ (2 tailed) ** = $p < 0.01$ level (2 tailed)
*** = $p < 0.001$ level (2 tailed)

The BNIS Speech and Language Subscale was correlated with most of the neuropsychological measures, except for Picture Completion, Grooved Pegs left hand, and the HCT. The strongest correlations were generally with neuropsychological measures sensitive to verbal abilities, such as the WAIS-R Verbal subtests, the WMS Logical Memory, and the COWAT.

The BNIS Visual Spatial and Visual Problem Solving Subscale was correlated strongest with neuropsychological measures sensitive to visual spatial and visual constructional abilities, such as the WAIS-R Performance subtests, Trail Making Test B, and the Grooved Pegboard left hand. In general, weaker correlations were found with neuropsychological measures sensitive to verbal abilities (e.g., the WAIS-R Verbal subtests and the COWAT), and the Grooved Pegboard right hand.

The BNIS Memory Subscale was correlated strongest with the delayed recall component of the WMS-R Logical Memory and Visual Reproduction subtests. The BNIS Affect Subscale was significantly correlated with most neuropsychological measures, except WMS-R Logical Memory I and II. Most of the significant correlations were within the $p < 0.001$ range, including all of the WAIS-R subtests, Trails A and B, and the WCST. In contrast to the other BNIS Total and Subscale measures, the BNIS Awareness Subscale was not significantly correlated with any of the neuropsychological measures. This finding with the Awareness Subscale is in contrast to the finding reported by Prigatano, Amin, and Rosenstein (1993), in which the Awareness Subscale was found to be significantly correlated with all neuropsychological measures.

Hypothesis Two.

A standard multiple regression was performed between the WAIS-R VIQ as the dependent variable and the BNIS Expression composite, Comprehension, Naming, Repetition, Reading, Writing composite, Spelling composite, and Arithmetic composite scores as the independent variables. Table 4 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F(8, 57) = 3.85$, $p < .001$.

Table 4

Standard Multiple Regression of BNIS Speech and Language Variables on WAIS-R VIQ

Variables	WAIS-R VIQ	Expres.	Compre	Naming	Repeti.	Read.	Writ.	Spell.	Arith.	*Beta*	Sr^2 *(unique)*
Expression	.31									.164	.02
Comprehension	.14	-.13								-.018	.00
Naming	.19	.02	.14							.142	.02
Repetition	.43	.10	.13	.12						.355	.11
Reading	.35	.29	.29	.21	.24					-.003	.00
Writing	.21	.23	.41	-.00	.23	.17				.062	.00
Spelling	.38	.30	.29	-.00	.09	.62	.11			.298	.05
Arithmetic	.09	-.03	.52	.01	.02	.12	.42	.23		.002	.00
Means	86.85	2.65	1.95	.88	1.71	.83	1.62	1.62	1.76	Unique variability = .20 ; shared variability = .15	
Standard Deviations	12.36	.59	.21	.33	.65	.38	.70	.72	.47		
$R = .592$ ($p < .001$)			$R^2 = .351$			Adjusted $R^2 = .260$					

The only two independent variables that contributed significantly to prediction of the WAIS-R VIQ score were Repetition ($sr^2 = .11$) and Spelling composite ($sr^2 = .05$). All independent variables in combination contributed 0.15 in shared variability. Altogether, 35 percent (26 % adjusted) of the variability in the WAIS-R VIQ scores was predicted by the scores on the eight independent variables.

Hypothesis Three.

A standard multiple regression was performed between the WAIS-R PIQ as the dependent variable and the BNIS Visual Object Recognition, Visual Scanning, Visual Sequencing, Pattern Copying, and Pattern Recognition scores as the independent variables. Table 5 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F(5, 63) = 2.48$, $p < .041$.

Table 5

Standard Multiple Regression of BNIS Visual Spatial & Visual Problem Solving Variables on WAIS-R PIQ

Variables	WAIS-R PIQ	VOR	Visual Scan	Visual Seq	Pattern Copy	Pattern Recog	*Beta*	Sr^2 (unique)
VOR	-.15						-.103	.01
Visual Scan	.20	-.13					.115	.01
Visual Seq	-.08	.06	.10				-.145	.02
Pattern Copy	.34	-.08	.26	.17			.325	.10
Pattern Recog	.08	.05	.18	.06	.09		.039	.00
Means	79.83	.99	1.1	.22	.72	.16	Unique variability = .14; shared variability = .02	
Standard Deviations	14.92	.12	.88	.42	.45	.37		
	$R = .406$ ($p <$ .041)		$R^2 = .164$		Adjusted $R^2 = .098$			

The only independent variable that contributed significantly to prediction of the WAIS-R PIQ score was Pattern Copying ($sr^2 = .10$). All independent variables in combination contributed 0.14 in shared variability. Altogether, only 16 percent (10 % adjusted) of the variability in the WAIS-R PIQ scores was predicted by the scores on the five independent variables.

Ecological Validity of the BNIS

Hypothesis Four.

Pearson correlation coefficients are presented in Table 6 between the FIM+FAM Total and Motor and Psychosocial/Cognitive Subtotals, and the BNIS Total and Subscales.

Table 6

Pearson Correlation Coefficients between the BNIS and the FIM+FAM

	BNIS Total	BNIS Speech	BNIS Orientation	BNIS Attention	BNIS Visual Spatial/Prob.	BNIS Memory	BNIS Affect	BNIS Awareness
FIM+FAM 1 – 16	.52 ***	.37 ***	.30 **	.31 **	.57 ***	.36 ***	.39 ***	.22 *
FIM+FAM 17 – 30	.66 ***	.49 ***	.47 ***	.37 ***	.47 ***	.55 ***	.52 ***	.44 ***
FIM+FAM Total 1 – 30	.65 ***	.47 ***	.42 ***	.37 ***	.57 ***	.49 ***	.50 ***	.36 ***

* = $p < 0.05$ (2 tailed)
** = $p < 0.01$ level (2 tailed)
*** = $p < 0.001$ level (2 tailed)

The FIM+FAM Psychosocial/Cognitive total score was significantly correlated (at the p<0.001 level) with all BNIS Subscales and the BNIS Total score, as was the FIM+FAM Total Score. The FIM+FAM Motor subscale total also exhibited a significant positive correlation with all BNIS measures, although the correlations were weaker as compared to the Psychosocial/Cognitive total for most BNIS Subscales. The BNIS Visual Spatial and Visual Problem Solving Subscale exhibited a stronger correlation with the FIM+FAM Motor score, as compared to the Psychosocial/Cognitive score.

Hypothesis Five.

A standard multiple regression was performed between the FIM+FAM Communications as the dependent variable and the BNIS Expression composite, Arithmetic composite, Comprehension, Naming, Repetition, Reading, Writing composite, and Spelling composite scores as the independent variables. Table 7 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F\,(8, 76) = 47.74, p < .001$.

Table 7

Standard Multiple Regression of BNIS Speech and Language Variables on FIM+FAM Communications

Variables	FIM+FAM Commun.	Expres.	Compre	Naming	Repeti.	Read.	Writ	Spell.	Arith.	*Beta*	Sr^2 *(unique)*
Expression	.54									.211	.03
Comprehension	.40	.43								.026	.00
Naming	.45	.27	.52							-.007	.00
Repetition	.49	.44	.25	.31						-.046	.01
Reading	.85	.30	.32	.47	.34					.113	.32
Writing	.44	.40	.52	.36	.14	.34				.688	.01
Spelling	.49	.35	.25	.37	.35	.37	.19			.103	.01
Arithmetic	.34	.34	.40	.35	.25	.23	.28	.49		.108	.00
Means	28.55	2.53	1.62	.87	1.61	5.71	1.55	1.46	1.62	Unique variability = .38 ; shared variability = .45	
Standard Deviations	5.09	.77	.60	.34	.76	1.11	.70	.76	.60		
$R = .913$ $(p < .001)$				$R^2 = .834$			Adjusted $R^2 = .817$				

The three independent variables that contributed significantly to prediction of the FIM+FAM Communication score were the Expression composite ($sr^2 = .03$), Repetition ($sr^2 = .01$) and Reading ($sr^2 = .32$). All independent variables in combination contributed 0.45 in shared variability. Altogether, 83 percent (82 % adjusted) of the variability in the FIM+FAM Communication score was predicted by the scores on the eight independent variables.

Hypothesis Six.

A standard multiple regression was performed between the FIM+FAM Orientation score as the dependent variable and the three BNIS Orientation items (i.e., Right – left orientation, Place, and Date) as the independent variables. Table 8 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F(3, 81) = 15.37, p < .001$.

Table 8

Standard Multiple Regression of BNIS Orientation Variables on FIM+FAM Orientation

Variables	FIM+FAM Orientation	Right - Left	Place	Date	*Beta*	Sr^2 (unique)
Right – Left	.07				-.049	.00
Place	.53	.12			.462	.20
Date	.41	.21	.27		.294	.08
Means	5.69	.92	.86	.49	Unique variability = .28 ; shared variability = .08	
Standard Deviations	1.73	.28	.35	.50		
	$R = .602$ ($p < .001$)		$R^2 = .363$	Adjusted $R^2 = .339$		

The independent variables that contributed significantly to prediction of the FIM+FAM Orientation score was Place ($sr^2 = .20$), and Date ($sr^2 = .08$). All independent variables in combination contributed 0.28 in shared variability. Altogether, 36 percent (34 % adjusted) of the variability in the FIM+FAM Orientation scores was predicted by the scores on the three independent variables.

Hypothesis Seven.

A standard multiple regression was performed between the FIM+FAM Attention score as the dependent variable and the three BNIS Attention items (i.e., Memory & Concentration, Digits Forward, and Digits Backward) as the independent variables. Table 9 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was not significantly different from zero, F (3, 81) = 2.01, $p < .118$.

Table 9

Standard Multiple Regression of BNIS Attention Variables on FIM+FAM Attention

Variables	FIM+FAM Attention	Memory & Concentration	Digits Forward	Digits Backward	*Beta*	Sr^2 (unique)
Memory & Conc	.22				.180	.03
Digits Forward	.07	.15			.011	.00
Digits Backward	.20	.26	.19		.148	.02
Means	5.61	.53	.79	.27	Unique variability = .05 ; shared variability = .02	
Standard Deviations	1.35	.50	.41	.45		
	$R = .263$ ($p < .118$)		$R^2 = .069$	Adjusted $R^2 = .035$		

None of the independent variables contributed significantly to prediction of the FIM+FAM Attention score. All independent variables in combination contributed 0.05 in shared variability. Altogether, 7 percent (3.5 % adjusted) of the variability in the FIM+FAM Attention scores was predicted by the scores on the three independent variables.

Hypothesis Eight.

A standard multiple regression was performed between FIM+FAM Problem Solving as the dependent variable and the BNIS Visual Scanning, Visual Sequencing, Pattern Copying, Pattern Recognition, and Awareness versus Performance scores as the independent variables. Table 10 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, F (5, 79) = 14.68, $p < .001$.

Table 10

Standard Multiple Regression of BNIS Visual Spatial & Visual Problem Solving

Variables on FIM+FAM Problem Solving

Variables	FIM+FAM Problem Solving	Visual Scan	Visual Seq	Pattern Copy	Pattern Recog	Aware vs. Perform	*Beta*	*Sr²* *(unique)*
Visual Scan	.15						.063	.00
Visual Seq	.46	.16					.364	.11
Pattern Copy	.35	.25	.29				.139	.02
Pattern Recog	.35	.16	.28	.24			.043	.00
Aware vs Perf	.53	-.02	.07	.19	.36		.460	.18
Means	4.16	1.35	.35	.61	.21	.39	Unique variability = .31 ; shared variability = .17	
Standard Deviations	1.74	.86	.48	.49	.41	.49		
	$R = .694$ ($p < .001$)		$R^2 = .482$		Adjusted $R^2 = .449$			

The two independent variables that contributed significantly to prediction of the FIM+FAM Problem Solving score were Visual Sequencing ($sr^2 = .11$), and Awareness versus Performance ($sr^2 = .18$). All independent variables in combination contributed 0.31 in shared variability. Altogether, 48 percent (45 % adjusted) of the variability in the FIM+FAM Problem Solving score was predicted by the scores on the five independent variables.

Hypothesis Nine.

A standard multiple regression was performed between the FIM+FAM Memory score as the dependent variable and the BNIS Learning & Memory and the three Delayed Recall items (i.e., House, Tree, and Flying) as the independent variables. Table 11 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F(4, 80) = 13.16$, $p < .001$.

Table 11

Standard Multiple Regression of BNIS Memory Variables on FIM+FAM Memory

Variables	FIM+FAM Memory	Learn & Memory	House	Tree	Flying	*Beta*	Sr^2 (unique)
Learn & Memory	.50					.241	.04
House	.49	.38				.187	.02
Tree	.47	.49	.64			.149	.01
Flying	.50	.50	.49	.35		.232	.03
Mean	4.71	1.48	.55	.54	.36	Unique variability = .10 ; shared variability = .30	
Standard Deviation	1.90	1.39	.50	.50	.48		
	$R = .630$ ($p < .001$)		$R^2 = .397$		Adjusted $R^2 = .367$		

The two independent variables that contributed significantly to prediction of the FIM+FAM Memory score were Learning & Memory ($sr^2 = .04$), and Delayed Recall – Flying ($sr^2 = .03$). All independent variables in combination contributed 0.10 in shared variability. Altogether, 40 percent (37 % adjusted) of the variability in the FIM+FAM Memory scores was predicted by the scores on the four independent variables.

Hypothesis Ten.

A standard multiple regression was performed between the FIM+FAM Emotional Status score as the dependent variable and the BNIS Affect items (i.e., Affect Expression, Perception of facial affect, Affect control, and Spontaneous Affect) as the independent variables. Table 12 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was not significantly different from zero, $F(4, 80) = 1.24$, $p < .302$.

Table 12

Standard Multiple Regression of BNIS Affect Variables on FIM+FAM Emotional Status

Variables	FIM+FAM Emotional Status	Affect Expression	Perception Facial Affect	Affect Control	Spontaneous Affect	*Beta*	*Sr*^{2} (unique)
Affect Expression	.03					-.052	.00
Perception of F A	.17	.33				.165	.02
Affect Control	.19	.15	.16			.166	.03
Spontaneous Affect	.04	.33	.18	.11		.006	.00
Mean	5.26	.71	.61	.91	.41	Unique variability = .05 ; shared variability = .01	
Standard Deviation	1.33	.46	.49	.29	.50		
	$R = .241$ ($p < .302$)		$R^2 = .058$		Adjusted $R^2 = .011$		

None of the independent variables contributed significantly to prediction of the FIM+FAM Emotional Stability score. All independent variables in combination contributed 0.05 in shared variability. Altogether, only 6 percent (1 % adjusted) of the variability in the FIM+FAM Memory scores was predicted by the scores on the four independent variables.

Hypothesis Eleven.

A standard multiple regression was performed between the FIM+FAM Psychosocial / Cognitive subtotal as the dependent variable and the BNIS Subscale totals (i.e., Speech and Language, Orientation, Attention/Concentration, Visual Spatial/Visual Problem Solving, Memory, Affect, and Awareness versus Performance) as the independent variables. Table 13 presents the correlations between the variables, the standardized regression coefficients (*Beta*), the semipartial correlations (sr^2), and R, R^2, and adjusted R^2. R for regression was significantly different from zero, $F(7, 77) = 12.68$, $p < .001$.

Table 13

Standard Multiple Regression of BNIS Subscale Variables on FIM+FAM Psychosocial-Cognition

Variables	FIM+FAM Psychosocial Cognition	Speech & Lang	Orient	Attention / Concen	Visual Spatial	Memory	Affect	Aware	*Beta*	*Sr^2 (unique)*
Speech & Lang	.49								.118	.01
Orientation	.47	.56							.220	.03
Attention/Conc	.37	.49	.44						-.028	.00
Visual Spatial	.47	.46	.35	.49					.131	.01
Memory	.55	.44	.50	.43	.42				.140	.01
Affect	.52	.42	.22	.35	.49	.38			.251	.04
Awareness	.44	.15	.09	.12	.15	.43	.26		.261	.05
Means	72.41	12.28	2.27	1.59	5.26	2.94	2.64	.39	Unique variability = .15; shared variability = .38	
Standard Deviations	15.40	3.20	.79	.93	1.70	2.30	1.13	.49		
$R = .732$ ($p < .001$)		$R^2 = .535$		Adjusted $R^2 = .493$						

The three independent variables that contributed significantly to prediction of the FIM+FAM Psychosocial-Cognition score were the Affect ($sr^2 = .04$), Orientation ($sr^2 = .03$) and Awareness versus Performance ($sr^2 = .05$). All independent variables in combination contributed 0.15 in shared variability. Altogether, 53.5 percent (49 % adjusted) of the variability in the FIM+FAM Psychosocial-Cognition score was predicted by the scores on the seven independent variables.

CHAPTER IV

DISCUSSION

The results of these analyses provide support for the construct-related validity of the BNIS through its relationship with neuropsychological measures. The ecological validity of the BNIS was supported through its relationship with the FIM+FAM. The discussion of these results is divided into twelve sections. The first three sections address the data analysis findings in relation to the construct-related validity of the BNIS as explored in the first three hypotheses. Sections four through eleven address the findings of data analyses associated with the hypotheses exploring the ecological validity of the BNIS. The final section outlines the primary contributions of this study and provides possible suggestions for future studies.

Hypothesis One

Hypothesis one stated that there would be a significant positive correlation between the BNIS Total and Subscale scores and selected neuropsychological test items. These results were expected based upon the correlation study findings reported by Prigatano, Amin, and Rosenstein (1993b). Hypothesis one was partially supported through the significant correlations between the BNIS Total score and all neuropsychological measures included in this study. Most of these correlations were significant at the $p < 0.001$ level, with the exception of the WMS-R Logical Memory I and Grooved Pegs-left which were both significant at the $p < 0.01$ level. This is consistent with the findings reported by Prigatano, Amin, and Rosenstein (1993b), although the strength of the correlations was not as high. The correlations reported by Prigatano, Amin, and

Rosenstein (1993b) ranged from 0.56 to 0.78, while the correlations in the present study ranged from 0.30 (HCT) to 0.57 (WAIS-R Digit Symbol). These results provide support for the construct-related validity of the BNIS as a summary measure of cognitive functioning.

In the study reported by Prigatano, Amin, and Rosenstein (1993b), all correlations between the neuropsychological measures and the BNIS Subscales (except Orientation with Trails B) were found to be significant at the $p < 0.001$ level. However, in the present study this same pattern was not repeated. This was most apparent with the BNIS Awareness Subscale, which was not significantly correlated with any of the neuropsychological measures. One possible explanation for this finding could be that the Awareness subtest is measuring something unique that is not captured by the neuropsychological measures. However, this conclusion would be inconsistent with the strong correlations reported by Prigatano, Amin, and Rosenstein (1993b).

Another possible explanation for the difference between studies could be that the patient groups between this study and the Prigatano studies were significantly different in composition. Although a direct analysis of the differences between these two groups would not be possible, a comparison of the presented demographic information might yield some useful insights. The patient sample described by Prigatano, Amin, and Rosenstein (1993b) had a mean age of 41.3 years (S.D. = 19.1) and a mean education of 12.6 years (S.D. = 2.9). The patient sample used in this study had a mean age of 36.5 (S.D. = 13.0) and a mean education of 11.7 (S.D. = 2.1). The mean BNIS score obtained by the Prigatano patient sample was 36.3 (S.D. = 7.6), which was consistent with the mean BNIS score of 36.3 (S.D. = 6.8) for the current patient sample. Twenty-nine percent of Prigatano's patient sample achieved a correct performance on the awareness question, while thirty-one percent of the patient sample in this study achieved a correct performance on the Awareness Subscale. These demographic variables and BNIS values between the two patient samples are seen to be generally consistent with each other.

However, the level of chronicity (i.e., time from date of injury to test date) appears to vary between the two samples. In Prigatano's study fifty-five percent of the patients had a chronicity of less than three months, twelve percent fell between three and twelve months, and thirty-three percent were above twelve months. In the patient sample used for this study only sixteen percent were tested within three months of injury, forty-eight percent were between three and twelve months, and thirty-six percent were above twelve months. Although the percentages between the two patient samples for chronicity above twelve months was similar, the percentages for chronicity below three months appear to be quite different. Considering the similarities between the two patient samples in regard to demographic variables and BNIS scores, it is likely that the difference between the two patient samples would be found in the neuropsychological test performance.

The majority of the patients in Prigatano's study were tested within three months of their trauma. It is possible that this resulted in a more uniform distribution of scores across all neuropsychological measures. On the other hand, this study had a relatively even distribution of patients across all levels of chronicity which may have produced a less even distribution of neuropsychological scores. Further studies would be required to explore this possibility. This could include an analysis of change in scores on the BNIS over time from injury. Also, the relationship between the BNIS scores and neuropsychological measures may vary significantly depending upon the length of time from injury to assessment. Exploration of these issues could help reveal specific patterns of performance over time on BNIS measures such as the Awareness item.

Although the remaining BNIS Subscales did not exhibit significant correlations with all neuropsychological measures, the pattern of correlations was consistent with expectations. The BNIS Speech and Language Subscale exhibited the strong significant correlations with neuropsychological measures sensitive to verbal abilities, such as the WAIS-R Verbal subtests, the WMS Logical Memory, and the COWAT. On the other hand, the correlation with Picture Completion, Grooved Pegs left hand, and the HCT failed to reach significance. The failure to reach significant correlation with Picture Completion and Grooved Pegs left hand was expected given the assumption that performance on these tasks would be dependent upon right hemisphere functioning, while speech and language functions would be dependent upon left hemisphere functioning.

The weak non-significant correlation between the BNIS Speech and Language and the HCT was also not unexpected. The HCT was included because of its sensitivity to brain damage regardless of its location. As such, the HCT was expected to correlate best with the BNIS Total score rather than any one BNIS Subscale. The HCT did, in fact, correlate weakly with almost all BNIS Subscales. However, the Affect Subscale exhibited a relatively strong correlation with the HCT. Both the HCT and the WCST exhibited the strongest correlations with the BNIS Affect Subscale. The BNIS Affect Subscale's pattern of strong correlations with most of the neuropsychological measures suggests that it could interpreted as an indicator of overall cognitive functioning, rather than specific to any one cognitive skill or ability.

The BNIS Visual Spatial and Visual Problem Solving Subscale exhibited the strongest correlations with neuropsychological measures sensitive to visual spatial and visual constructional abilities, such as the WAIS-R Performance subtests, Trail Making Test B, and the Grooved Pegboard left hand. In general, weaker correlations were found with neuropsychological measures sensitive to verbal abilities (e.g., the WAIS-R Verbal subtests and the COWAT), and the Grooved Pegboard right hand which was assumed to be related to the integrity of the left cerebral hemisphere. These findings were expected given the assumption that performance on the BNIS Visual Spatial and Visual Problem Solving Subscale tasks would be dependent upon right hemisphere functioning.

The BNIS Memory Subscale exhibited the strongest correlations with the delayed recall component of the WMS-R Logical Memory and Visual Reproduction subtests, and the WAIS-R Digit Symbol subtest. The strong correlation between the BNIS Memory Subscale and the WAIS-R Digit Symbol is not surprising given the similarity between the WAIS-R Digit Symbol subtest and the number-symbol learning task in the BNIS Memory Subscale. This pattern of correlations indicates that the BNIS Memory Subscale appears to be assessing memory functioning as determined through the neuropsychological criterion measures.

In regard to the neuropsychological measures, the highest correlations were typically found with those BNIS Subscales that were developed to assess similar abilities. For example, the WAIS-R Verbal IQ subtests and the COWAT exhibited the highest correlations with the BNIS Speech and Language Subscale. Similarly, the WAIS-R Performance IQ subtests generally exhibited the highest correlations with the BNIS Visual Spatial and Visual Problem Solving Subscale. The WMS-R delayed recall tasks exhibited the highest correlations with the BNIS Memory Subscale. The right hand Grooved Pegboard Test exhibited the strongest correlation with the BNIS Speech and Language Subscale and the Memory Subscale, and exhibited only a weak correlation with the BNIS Visual Spatial and Visual Problem Solving Subscale. Conversely, the left hand Grooved Pegboard Test exhibited the strongest correlation the BNIS Visual Spatial and

Visual Problem Solving Subscale and the Affect Subscale, and a non-significant correlation with the Speech and Language Subscale.

In summary, these results support the hypothesis that there will be a significant positive correlation between neuropsychological measures and the BNIS Total and Subscale scores. Further to this, is the finding that the BNIS Speech and Language, Visual Spatial and Visual Problem Solving, and Memory Subscales exhibited the highest correlations with those neuropsychological measures that were designed to measure similar skill and ability areas. Therefore, these results provide support for the construct-related validity of the BNIS as determined through its significant correlation with the neuropsychological criterion measures.

Hypothesis Two

Hypothesis two stated that the BNIS Speech and Language Subscale items would predict performance on the WAIS-R Verbal IQ measure. The significant predictive relationship that was found in this study supports the construct-related validity of the BNIS Speech and Language Subscale. Univariate correlation analyses revealed that the BNIS Expression composite, Repetition, Reading, Writing composite, and the Spelling composite were all significantly correlated with the WAIS-R Verbal IQ measure. In contrast, the BNIS Comprehension, Naming, and Arithmetic composite scores were not significantly correlated with the WAIS-R Verbal IQ measure.

The moderate strength of this predictive relationship was not unexpected given the differences between the two measures. On the one hand, the WAIS-R Verbal IQ provides a summary score of a person's overall verbal abilities, based upon a summary of performance across a variety of disparate but related verbal abilities, including verbal comprehension and expression. Performance on the WAIS-R Verbal measures will also be significantly affected by many other factors, such as attention and concentration, ability to recall previously learned information, higher-order reasoning, level of education, and environmental deprivation. On the other hand, the BNIS Speech and Language items are single questions designed to provide a screening evaluation of the integrity of a person's basic speech and language capacities.

The WAIS-R is capable of providing a more fine-grained analysis of an individual's verbal abilities, and produces an indicator of a person's verbal abilities at any level, from severely impaired through to very superior performance. The BNIS is intended to provide a simple yes/no indicator of a person's ability to perform a given task. The intention behind the inclusion of the BNIS items is that any neurologically unimpaired individual would be able to complete each item. Therefore, a failure on any item is worthy of further examination to determine the origins of that failure. The BNIS was designed to provide a gross assessment of an individual's abilities, based upon a minimum amount of data in order to keep the assessment brief. The BNIS was not designed to replace the need for a more comprehensive assessment of a person's cognitive abilities, but to allow for the identification and tracking of possible cognitive weaknesses. The significant predictive relationship between the BNIS Speech and Language items and the WAIS-R Verbal IQ indicates that the BNIS Speech and Language items are evaluating a significant portion of the same underlying construct. Therefore, it was felt that the results of this analysis provide support for the construct-related validity of the BNIS.

Hypothesis Three

Hypothesis three stated that the BNIS Visual Spatial and Visual Problem Solving Subscale items would predict performance on the WAIS-R Performance IQ measure. The significant predictive relationship that was found in this part of the study supports the construct-related validity of the BNIS Visual Spatial and Visual Problem Solving Subscale. Univariate correlation analyses revealed that the BNIS Visual Scanning and Pattern Copying were significantly correlated with the WAIS-R Performance IQ measure. In contrast, the BNIS Visual Object Recognition, Visual Sequencing, and Pattern Recognition were not significantly correlated with the WAIS-R Performance IQ measure.

The significant predictive relationship between the BNIS items and the WAIS-R Performance IQ indicates that the BNIS items are assessing a portion of a similar underlying construct. However, this relationship was considerably weaker than was found between the BNIS Speech and Language scores and WAIS-R Verbal IQ. The same issues that were discussed in hypothesis three regarding the WAIS-R and the BNIS would be relevant here. The relatively weak predictive relationship with the WAIS-R Performance IQ would suggest that there are other significant factors are likely to be underlying the BNIS Visual Spatial and Visual Problem Solving items. For example, the Visual Sequencing item has been observed during clinical work to be best described as a measure of working memory, rather than as a perceptual organizational task. Further studies of the BNIS items would be required to help elucidate the underlying cognitive factors assessed in the BNIS Visual Spatial and Visual Problem Solving items. In conclusion, the results from this study support the hypothesis that the BNIS Visual Spatial and Visual Problem Solving items are assessing a visual perceptual organizational cognitive construct, although the relatively weak relationship indicates that a large portion of the variance was unaccounted for in this equation.

<u>Hypothesis Four</u>

Hypothesis four stated that the discharge BNIS Total and Subscale scores would exhibit a significant positive correlation with the discharge Total FIM+FAM score, as well as the FIM+FAM Motor and Psychosocial/Cognitive Subtotal scores. The significant positive relationships that were found in this part of the study supports the ecological validity of the BNIS. Also consistent with the hypothesis was the finding that most of the BNIS scores exhibited a stronger positive correlation with the FIM+FAM Psychosocial/Cognitive Subtotal scores as compared to the FIM+FAM Motor Subtotal scores. The only exception to this pattern was found with the BNIS Visual Spatial and Visual Problem Solving, in which the relationship with the FIM+FAM Motor Subtotal was stronger.

Although the relationship between the BNIS Visual Spatial and Visual Problem Solving and the FIM+FAM Subtotal scores was not predicted, similar findings have been reported in studies comparing neuropsychological test performance with activities of daily living (Shaw & Groom, 1996). For example, Richardson, Nadler, and Malloy (1995) examined the predictive relationship between neuropsychological test performance and the ability to perform basic functional skills and everyday living skills in a geriatric population. Evaluation of the activities of daily living (ADL) was assessed through standardized measures of hygiene/self-care, safety, cooking and meal preparation, money management, and community access. The results of their study were that the measures assessing visual spatial abilities exhibited the strongest correlations with most of the ADL measures and the highest Positive Predictive Power. Their study helps clarify the significant role that visual spatial abilities play in an individual's ability to perform basic functional and everyday skills.

The importance of visual perceptual skills to the performance of basic functional skills and outcome was found in a rehabilitation study in a population of stroke patients receiving physical therapies. Finlayson, Gowland, and Basmajian (1986; as cited in Acker, 1990) in a study of 29 stroke patients who were assigned to two treatment groups, evaluated change from pre- to post-treatment with a variety of measures. They stressed the importance of disturbed perceptual functioning as a limiting factor in stroke recovery. It was found that the best outcome predictors were those that evaluate immediate adaptive skill and visual-perceptual-motor skills.

The stronger relationship between the BNIS Visual Spatial and Visual Problem Solving with the FIM+FAM Motor Subscale scores would be consistent with the findings reported by Richardson, Nadler, and Malloy (1995). The FIM+FAM Motor Subscale includes an evaluation of self-care items such as eating, grooming, bathing, and dressing, as well as mobility items such as walking/wheelchair locomotion and community mobility. The results of this study indicate that visual spatial abilities play a significant role in the ability to perform these basic ADL's.

Hypothesis Five

Hypothesis five stated that the discharge BNIS Speech and Language Subscale scores would predict performance on the FIM+FAM Communications subtotal. Univariate correlation analyses revealed that all of the BNIS Speech and Language variables were significantly correlated with the FIM+FAM Communications subtotal. The significant predictive relationship that was found in this part of the study supports the ecological validity of the BNIS. The strong predictive relationship indicates that the BNIS Speech and Language scale provides a reasonably accurate reflection of how effectively an individual will be able to communicate in daily activities.

Hypothesis Six

Hypothesis six stated that the three discharge BNIS Orientation Subscale scores would predict performance on FIM+FAM Orientation. Univariate correlation analyses revealed that the BNIS Place and Date variables were significantly correlated with the FIM+FAM Orientation item, although the BNIS Right-Left item was not significantly correlated. The significant predictive relationship that was found in this part of the study supports the ecological validity of the BNIS.

Due to the highly variable nature of a patient's orientation to date and place at any given time it was not unexpected that the predictive relationship was not stronger. Their orientation can be affected not only by their own internal states (e.g., clearing from delirium, densely amnestic), but also by a variety of external stimuli (e.g., calendars, clocks) and cueing. The FIM+FAM Orientation item provides an overall evaluation of the patient's orientation to person, place, time, and situation based upon their contact with many different therapists over a period of time. On the other hand, the BNIS orientation items are a single point in time assessment of the patient's ability to accurately state the date and place. If they are inaccurate by any margin on the BNIS Date item they do not obtain a point for that item.

The FIM+FAM Orientation item attempts to address an individual's orientation to person, place, and time. The BNIS Right-Left orientation item requires the patient to follow a verbal direction to place their left hand to their right ear. Although this item attempts to address issues of orientation to body space, other factors are likely to play a significant role, including the patients ability to accurately comprehend the verbal instructions. These issues are not addressed in the scoring of the FIM+FAM Orientation item. Because of these differences in assessment techniques between the BNIS and FIM+FAM, it would probably be difficult to achieve a stronger predictive relationship. Therefore, it is indicated that the ecological validity of the BNIS Orientation was supported.

Hypothesis Seven

Hypothesis Seven stated that the discharge BNIS Attention/Concentration Subscale scores would predict performance on the discharge FIM+FAM Attention measure. Univariate correlation analyses revealed that the BNIS Memory/Concentration and Digits-Backward variables were significantly correlated with the FIM+FAM Attention measure, but that the BNIS Digits-Forward variable was not significantly correlated. The significant univariate correlations were relatively low, and reached significance only at the 0.05 level. The failure to achieve a significant predictive relationship did not support the ecological validity of the BNIS Attention items.

The FIM+FAM Attention item attempts to assess the length of time a patient is able to concentrate on a task, with the issues such as distractibility, level of responsiveness, and difficulty of the task being taken into account. The various levels assigned to a patient are determined primarily by the overall length of time they attend to a given task, with some focus on the percent of that time that is directed at the task at hand. A patient would receive the lowest score on the FIM+FAM Attention item if they were unable to attend to a task for less than one minute. They would receive the second lowest score if they attend for only one to four minutes with attention to the task at hand for at least 25 to 49 percent of that time. The criteria established for the FIM+FAM Attention measure appears to be largely dependent upon sustained attention and/or vigilance.

In comparison, the BNIS Attention items are unlikely to require more than approximately one minute to complete any one of the three items. The BNIS Attention items require the ability to concentrate upon and maintain verbally presented material, as well as intact working memory capacities. A possible reason for the failure to find a significant predictive relationship was a result of these differences between the FIM+FAM Attention scoring criteria and the required cognitive capacities for successful completion of the BNIS Attention items. Further studies would be required to determine if there are ecologically valid indicators of attention capacities that appear to be similar to the cognitive capacities required for the BNIS Attention items. Performing predictive analyses using that criteria could provide data to support the ecological validity of the BNIS Attention items.

Hypothesis Eight

Hypothesis eight stated that the discharge BNIS Visual Scanning, Visual Sequencing, Pattern Copying, Pattern Recognition, and Awareness versus Performance scores would predict performance on discharge FIM+FAM Problem Solving item. Univariate correlation analyses revealed that the all BNIS variables except Visual Scanning were significantly correlated with the FIM+FAM Problem Solving item. The significant predictive relationship that was found in this part of the study supports the ecological validity of the BNIS.

The general scoring criteria for the FIM+FAM Problem Solving item attempts to assess a patient's skills related to solving problems of daily living through prudent, safe, and reasonable decisions regarding financial, social, and personal affairs. Almost half of the variability in the FIM+FAM Problem Solving scores was predicted by the BNIS variables. This indicates that a patient's performance on the BNIS variables will provide a reasonable prediction of their ability to engage in daily living problem solving skills as described by the FIM+FAM. The significant contribution of the BNIS Visual Sequencing and the Awareness versus Performance items are particularly well adapted to predicting every day problem solving abilities.

The BNIS Awareness versus Performance item's strong correlation and predictive relationship with the FIM+FAM Problem Solving measure supports its use as an indicator of *functional* higher order executive functioning. As discussed in hypothesis one of this study, the Awareness versus Performance item did not correlate significantly with any of the neuropsychological measures. Neuropsychological measures are typically designed to define areas of cognitive strength or impairment. Whether or not a cognitive impairment manifests itself into a functional disability or handicap can be determined by numerous factors. These factors could include a person's social conditions, expectations, and demands, and the availability of compensatory aids and strategies. The significant correlation and predictive relationship between the Awareness versus Performance item and the FIM+FAM Problem Solving indicates that the Awareness item is more reflective of *functional* ability rather than a measure of cognitive impairment.

The Visual Sequencing item requires the patient to visually scan five rows of nine numbers each and decide which two rows have the same set of numbers. The numbers within each row are not in the same order. Successful completion of this task is dependent upon intact working memory capacities. Imaging studies have demonstrated that working memory is largely dependent upon intact prefrontal cortices (e.g., Beardsley, 1997; Golman-Rakic, 1993). The prefrontal cortex is also identified as being critical to problem solving abilities through control of various executive functions. The nature of the scoring criteria for the FIM+FAM Problem Solving item suggests that it is attempting to also assess the same executive level processes. The strong correlation and predictive relationship between the BNIS Visual Sequencing and the FIM+FAM Problem Solving indicates that a similar underlying construct is being assessed.

<u>Hypothesis Nine</u>

Hypothesis nine stated that the discharge BNIS Memory items would predict performance on discharge FIM+FAM Memory item. Univariate correlation analyses revealed that the all BNIS variables were significantly correlated with the FIM+FAM Memory item. The significant predictive relationship that was found in this part of the study supports the ecological validity of the BNIS. Although the Learning and Memory item and the Delayed Recall-Flying item were identified as providing a significant independent contribution, all BNIS memory variables exhibited a relatively strong and highly significant correlation with each other.

The FIM+FAM Memory item attempts to evaluate the patient's ability to store and retrieve verbal and visual information, particularly as it pertains to their ability to perform daily activities. A relatively high score can be achieved on this measure through the use of various memory aids and compensatory strategies. For example, a score of "modified independence" is indicated if that person has some difficulty with recalling daily routines or people, but is utilizing self-initiated environmental cues, prompts, or aids. This item is focused upon the functional capabilities of the patient's recall, rather than any possible underlying memory impairment. The BNIS Memory items are designed to assess the patient's learning and memory abilities without the benefit of compensatory strategies during recall. The difference between the FIM+FAM Memory assessment of *functional capacity* as compared to the BNIS Memory assessment of memory *impairment* would account for at least some of the unexplained variability in the FIM+FAM Memory scores.

Hypothesis Ten

Hypothesis ten stated that the discharge BNIS Affect items would predict performance on discharge FIM+FAM Emotional Status item. Univariate correlation analyses revealed that only the BNIS Affect Control variable was significantly correlated with the FIM+FAM Emotional Status item. A significant predictive relationship was not found in this part of the study, and therefore does not support the ecological validity of the BNIS Affect items.

The FIM+FAM Emotional Status item evaluates the extent to which depression, anxiety, frustration, or agitation interferes with their ability to interact with others in their daily life. The significant correlation between this item and the BNIS Affect Control item indicates that these two are assessing a similar underlying construct. However, the remaining BNIS items appear to be assessing a different construct and therefore, were not significantly related to or predictive of the FIM+FAM Emotional Status item.

Hypothesis Eleven

Hypothesis eleven stated that the discharge BNIS Subscale Total scores would predict performance on discharge FIM+FAM Psychosocial-Cognition Subtotal. Univariate correlation analyses revealed that the all BNIS variables were significantly correlated with the FIM+FAM Psychosocial-Cognition Subtotal. The significant predictive relationship that was found in this part of the study supports the ecological validity of the BNIS.

The FIM+FAM Psychosocial-Cognition Subtotal was presented in this study as an overall measure of a patient's daily functional cognitive capabilities. Ratings on this scale were dependent upon their interactions during assessments and treatment sessions with the professional staff. As described earlier, these FIM+FAM ratings were designed to be an assessment of their presentation in these interactions, rather than an attempt to delineate any potential underlying impairment. Use of compensatory strategies and aids, along with the highly structured inpatient hospital environment could allow any given patient to present as relatively functionally intact. However, that same patient could exhibit significant cognitive impairment on psychometric assessment. Even given the potential discrepancy between impairment and functional capabilities, the BNIS variables were still able to predict about half of the variance in the FIM+FAM scores. Therefore, it is argued that strong evidence was found for the ecological validity of the BNIS.

Summary and Conclusion

The results of this study provide evidence to support both the construct-related and the ecological validity of the BNI Screen for Higher Cerebral Functions. Support for the validity of the BNIS allows the users of the instrument to have greater confidence in the findings produced when it is administered to any given patient. Clinicians can utilize the BNIS as a screening measure to delineate specific areas of cognitive strengths and weakness, as well as overall cognitive integrity. This is not to imply that the BNIS can or should take the place of a more comprehensive cognitive assessment. Rather, it can provide a brief screening examination useful in the initial diagnosis of a patient's cognitive abilities, and follow-up monitoring of a patient's progress through various treatment programs.

Anastasi (1988) indicated that construct-related validity is derived from established inter-relationships among behavioral measures, and that this is gathered through the gradual accumulation of information from a variety of sources. The construct-related validity of the BNIS was supported through the significant relationships between the BNIS items and various neuropsychological measures. These significant relationships indicate that clinicians and researchers can utilize this measure with reasonable confidence that it is actually measuring what it is purported to measure. The findings in this study are generally consistent with the results of the BNIS validity studies conducted by Prigatano and others in the development of the BNIS. Therefore, the convergence of results between the results of this study and those of Prigatano's studies would argue in favor of the construct-related validity of the BNIS.

The significant predictive relationships between the BNIS and FIM+FAM provided support for the ecological validity of the BNIS. In fact, some items were better predictors of functional ability than level of cognitive impairment. For example, the Awareness versus Performance item was found to be a good predictor of functional problem solving abilities, but was not significantly related to any of the measures of cognitive impairment. This makes this item particularly useful as an indicator of functional ability. Based upon the results of this study, a Neuropsychologist who had administered the BNIS to a patient would be able to make statements about a person's functional abilities with some confidence about the validity of those statements.

Further study of the BNIS would be recommended to explore its psychometric properties and clinical utility. A factor analytic study could provide evidence to support the factor structure of the BNIS Subscales. For example, in this study the Visual Object Recognition item, which is part of the Visual Spatial and Visual Problem Solving Subscale, did not correlate significantly with the WAIS-R PIQ. One possibility is simply that the WAIS-R PIQ is assessing different underlying abilities that are not captured in the Visual Object Recognition item. On the other hand, the Visual Object Recognition item is similar to tasks that are often included in aphasia batteries, and may be better included as a measure of speech and language functioning. Factor analytic studies could help provide information to examine the present factor structure of the BNIS.

A shortcoming of this study was that a large enough pool of subjects was not available to conduct a cross-validation study on the regression equations. A new sample of subjects could be selected for a cross-validation study to evaluate the original formula as further data is gathered from the patients at the hospital. This would allow a comparison of subject samples gathered from the same population. It would also be important to conduct cross-validation studies with these regression equations on subject samples gathered from different populations. This information could help determine if the BNIS exhibits substantially different characteristics with different subject populations. For example, the subject sample used in this study presented with moderate to severe brain injuries. With that subject sample the BNIS Speech and Language Variables predicted approximately one third of the variability in the WAIS-R VIQ. Factors that are unique to that patient population (e.g., significant attention and concentration problems) are likely to have had an impact on the other variables. A sample of individuals with mild head injuries is not likely to exhibit the same degree of difficulty with other factors, such as attention and concentration. Therefore, their performance on the BNIS Speech and Language items may be more predictive of their performance on the WAIS-R VIQ.

Another issue would be in regard to the use of the FIM+FAM as a measure of functional ability to perform activities of daily living. Acker (1990) provides a list of 57 functional assessment tools, which is not an exhaustive list of available instruments. It is a certainty that there would be many new measures that could be added since that list was compiled considering the increasing interest in issues of ecological validity and assessing functional abilities. For example, the FIM+FAM was not included on that list. Acker (1990) stated that functional assessments instruments will typically provide an evaluation of a number of personal and behavioral variables, including self-care and ADL, communication, mobility, psychosocial adjustment, education/employment, and other activities (e.g., money management, leisure activities).

An argument in favor of the FIM+FAM for use in this study is that it was designed specifically for use with patients who have suffered a brain injury and are in rehabilitation. The extensive base of research conducted on the FIM+FAM also allowed the concerns expressed by Franzen and Wilhelm (1996) about functional assessment instruments to be addressed. Although the use of the FIM+FAM as a measure of functional ability was defensible, further studies should be conducted to explore the relationship of the BNIS to other well-designed measures of functional outcome. The relationship between the BNIS and functional assessment instruments could also be examined throughout active rehabilitation. This could help to illuminate the relationship between the level of cognitive impairment and the level of functional abilities throughout the rehabilitation process. Another important set of studies could explore the relationship between the BNIS and follow-up outcome after discharge from an active rehabilitation program. This could help provide information on the relationship between levels of cognitive impairment and how those might relate to a patient's ability to reintegrate into his or her environment.

Overall, the results found in this study provide further support for the use of the BNIS in populations of moderate to severely brain injured patients in active rehabilitation. The BNIS, as with any screening instrument, is not designed to replace the need for a more comprehensive assessment. A screening instrument should provide a relatively brief, efficient, and cost-effective means of determining a person's abilities and needs. Any psychometric instrument requires careful examination of its reliability and validity in order to ensure that the results of the assessment provide accurate statements about the client. The BNIS thus far has proven to be a reliable and valid assessment measure with the populations that have been studied to date. There is a clear need to further study its psychometric properties with different populations, both as a clinical and a research instrument.

REFERENCES

Acker, M. B. (1990). A review of the ecological validity of neuropsychological tests. In D. E. Tupper & K. D. Cicerone (Eds.), *The Neuropsychology of Everyday Life: Assessment and Basic Competencies* (pp. 19-56). Boston: Kluwer Academic Publishers.

Alexander, M. P. (1994). Stroke rehabilitation outcome: A potential use of predictive variables to establish levels of care. *Stroke*, *25*, 128-134.

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

Anastasi, A. (1988). *Psychological Testing*. (6th ed.). New York: Macmillan Publishing Co.

Andiel, C. (1995). Rasch analysis: A description of the model and related issues. *Canadian Journal of Rehabilitation*, *9*, 17-25.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry*, *40*, 569-573.

Bartko, J. J. & Carpenter, W. T. (1976). On the methods and theory of reliability. *The Journal of Nervous and Mental Disease*, *163*, 307-317.

Batavia, A. I. (1988). *The payment of medical rehabilitation services: Current mechanisms and potential models*. Chicago: American Hospital Association.

Beardsley, T. (1997, August). The machinery of thought. *Scientific American*, 78-83.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15-22.

Blessed. G., Tomlinson, B., & Roth, M. (1968). The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly patients. *British Journal of Psychiatry*, *114*, 797-811.

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, *14*, 119-186.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing the interrater reliability of specific items in a given inventory: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127-137.

Committee to Develop Standards for Educational and Psychological Testing. (1985). *Standards for Educational and Psychological Testing*. Washington, DC; American Psychological Association.

Cowen, T. D., Meythaler, J. M., DeVito, M. J., Ivie, C. S., Lebow, J., & Novack, T. A. (1995). Influence of early variables in traumatic brain injury on Functional Independence Measure scores and rehabilitation length of stay and charges. *Archives of Physical Medicine and Rehabilitation*, *76*, 797-803.

Crewe, N. M. & Dijkers, M. (1995). Functional assessment. In L. A. Cushman & M. J. Scherer (Eds.), *Psychological Assessment in Medical Rehabilitation* (pp. 101-144). Washington, DC: American Psychological Association.

Cushman, L. A. & Scherer, M. J.(Eds.), (1995). *Psychological Assessment in Medical Rehabilitation*. Washington, DC: American Psychological Association.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult Version*. Toronto: The Psychological Corporation.

Emhoff, T. A., McCarthy, M., Cushman, M., Garb, J. L., Valenziano, C. (1991). Functional scoring of multi-trauma patients: Who ends up where? *The Journal of Trauma*, *31*, 1227-1232.

Engelhart, C., Eisenstein, N., & Meininger, J. (1994). Psychometric properties of the Neurobehavioral Cognitive Status Exam. *The Clinical Neuropsychologist*, *8*, 405-415.

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*, 93-102.

Finlayson, M. A. J., Gowland, C., & Basmajian, J. V. (1986, February). Neuropsychological predictors of treatment response following stroke. Paper presented at the annual meeting of the International Neuropsychological Society, Denver.

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

Franzen, M. D. & Wilhelm, K. L. (1996). Conceptual foundations of ecological validity in neuropsychological assessment. In R. J. Sbordone & C. J. Long (Eds.), *Ecological Validity of Neuropsychological Testing* (pp. 91-112). Delray Beach, FL: St. Lucie Press.

Goldman-Rakic, P. S. (1993). Specification of higher cortical functions. *Journal of head Trauma Rehabilitation*, *8*, 13-23.

Goldstein, G. (1996). Functional considerations in neuropsychology. In R. J. Sbordone & C. J. Long (Eds.), *Ecological Validity of Neuropsychological Testing* (pp. 75-90). Delray Beach, FL: St. Lucie Press.

Granger, C. V., Cotter, A. C., Hamilton, B. B., & Fielder, R. C. (1993). Functional assessment scales: A study of persons after stroke. *Archives of Physical Medicine and Rehabilitation, 74*, 133-138.

Granger, C. V., Hamilton, B. B., & Fielder, R. C. (1992). Discharge outcome after stroke rehabilitation. *Stroke, 23*, 978-982.

Guide for the Uniform Data Set for Medical Rehabilitation (Adult FIM), Version 4.0. (1993). Buffalo, NY: State University of New York at Buffalo.

Hall, K. M., Hamilton, B. B., Gordon, W. A., & Zasler, N. D. (1993). Characteristics and comparisons of functional assessment indices: Disability Rating Scale, Functional Independence Measure, and Functional Assessment Measure. *Journal of Head Trauma Rehabilitation, 8(2)*, 60-74.

Hall, K. M., & Johnston, M. V. (1994). Outcomes evaluation in TBI rehabilitation: Part II: Measurement tools for a nationwide data system. *Archives of Physical Medicine and Rehabilitation, 75*, SC10-SC18.

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

Hamilton, B. B., Laughlin, J. A., Granger, C. V., & Kayton, R. M. (1991). Interrater agreement of the seven level Functional Independence Measure (FIM) [Abstract]. *Archives of Physical Medicine and Rehabilitation, 72*, 790.

Haynes, R. B. (1981). How to read clinical journals: II. To learn about a diagnostic test. *Canadian Medical Association Journal, 124*, 703-710.

Heaton, R. K. (1981). *A Manual for the Wisconsin Card Sorting Test*. Odessa, FL.: Psychological Assessment Resources, Inc.

Heinemann, A. W., Linacre, J. M., Wright, B. D., Hamilton, B. B., & Granger, C. (1993). Relationships between impairment and physical disability as measured by the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation*, *74*, 566-573.

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.

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

Linacre, J. M., Heinemann, A. W., Wright, B. D., Granger, C. V., & Hamilton, B. B. (1994). The structure and stability of the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation*, *75*, 127-132.

Malloy, P. F., Cummings, J. L., Coffey, C. E., Duffy, J., Fink, M., Lauterbach, E. C., Lovell, M., Royall, D., Salloway, S. (1997). Cognitive screening instruments in neuropsychiatry: A report of the committee on research of the American Neuropsychiatric Association. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *9*, 189-197.

Mattis, S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In L. Bellak & T. B. Karasu (Eds.), *Geriatric Psychiatry: A Handbook for Psychiatrists and Primary Care Physicians* (pp. 77-121). New York: Grune & Stratton.

Norusis, M. J. (1997). *SPSS for Windows Base System User's Guide Release 7.5*. Chicago: SPSS Incorporated.

Northern California Behavioral Group (1988). *Manual for the Neurobehavioral Cognitive Status Examination*. Fairfax, CA: Northern California Behavioral Group, Inc.

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

Prigatano, G. P. (1991). BNI Screen for Higher Cerebral Functions: Rationale and initial validation. *BNI Quarterly, 7*, 2-9.

Prigatano, G. P. (1994). Assessment of Alzheimer's patients with the BNI Screen for Higher Cerebral Functions. *BNI Quarterly, 10*, 30-35.

Prigatano, G. P., Altman, I. M., & O'Brien, K. P. (1990). Behavioral limitations that brain-injured patients tend to underestimate. *The Clinical Neuropsychologist, 4*, 163-176.

Prigatano, G. P., Amin, K., & Rosenstein, L. D. (1993a). *Manual for the BNI Screen for Higher Cerebral Functions*. Barrow Neurological Institute: Phoenix, AZ.

Prigatano, G. P., Amin, K., & Rosenstein, L. D. (1993b). Validity studies on the BNI Screen for Higher Cerebral Functions. *BNI Quarterly, 9*, 2-9.

Prigatano, G. P., & Leathem, J. M. (1993). Awareness of behavioral limitations after traumatic brain injury: A cross-cultural study of New Zealand Maoris and non-Maoris. *The Clinical Neuropsychologist, 7*, 123-135.

Prigatano, G. P., & Smason, I. (1994). Assessment of Alzheimer's patients with the BNI Screen for Higher Cerebral Functions. *BNI Quarterly, 10*, 30-35.

Reid, J. C., & Bonwich, E. B. (1989). An investigation of the factor structure of the National Uniform Data System Functional Independence Measure [Abstract]. *Archives of Physical Medicine and Rehabilitation, 70*(Suppl. A), 31.

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

Richardson, E. D., Nadler, J. D., & Malloy, P. F. (1995). Neuropsychologic prediction of performance measures of daily living skills in geriatric patients.

Rosenstein, L. D., Prigatano, G. P., & Amin, K. (1992). Reliability studies for the BNI Screen for Higher Cerebral Functions. *BNI Quarterly*, *8*, 24-28.

Sbordone, R. J. & Long, C. J. (Eds.), (1996). *Ecological Validity of Neuropsychological Testing*. Delray Beach, FL: St. Lucie Press.

Segal, M. E., & Schall, R. R. (1994). Determining functional/health status and its relation to disability in stroke survivors. *Stroke*, *25*, 2391-2397.

Shaw, T. G. & Groom, K. N. (1996, November/December). Functional treatment recommendations for patients with right hemisphere, nonverbal deficits. *The Journal of Cognitive Rehabilitation*, 6-11.

Smith, A. (1976). *Symbol Digit Modalities Test*. Los Angeles, CA: Western Psychological Services.

Smith, P. M., Illig, S. B., Fielder, R. C., Hamilton, B. B., Ottenbacher, K. J. (1996). Intermodal agreement of follow-up telephone assessment using the Functional Independence Measure in patients with stroke. *Archives of Physical Medicine and Rehabilitation*, *77*, 431-439.

Tabachnick, B. G. & Fidell, L. S. (1989). *Using Multivariate Statistics* (2nd Ed.). New York, NY: Harper & Row Publishers.

Tupper, D. E. & Cicerone, K. D. (1990), *The Neuropsychology of Everyday Life: Assessment and Basic Competencies*. Boston: Kluwer Academic Publishers.

Wagner, M. T., Nayak, M., & Fink, C. (1995). Bedside screening of neurocognitive function. In L. A. Cushman & M. J. Scherer (Eds.), *Psychological Assessment in Medical Rehabilitation* (pp. 145-198). Washington, DC: American Psychological Association.

Wass, P. J. (1996a). *Neuropsychological screening measures: Comparison of the Neurobehavioral Cognitive Status Examination and the BNI Screen for Higher Cerebral Functions*. Unpublished manuscript, University of Windsor.

Wass, P. J. (1996b) *Inter-Rater reliability of the Functional Assessment Measure within a brain injury rehabilitation program*. Unpublished manuscript, University of Windsor.

Webb, C. R., Wrigley, M., Yoels, W., Fine, P. R. (1995). Explaining quality of life for persons with traumatic brain injuries 2 years after injury. *Archives of Physical Medicine and Rehabilitation*, *76*, 1113-1119.

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

Wechsler, D. (1987). *The Wechsler Memory Scale - Revised*. San Antonio: The Psychological Corporation.

Wilkerson, D. L., Batavia, A. I., & DeJong, G. (1992). Use of functional status measures for payment of medical rehabilitation services. *Archives of Physical Medicine and Rehabilitation*, *73*, 111-120.

Wilson, D. B., Houle, D. M., & Keith, R. A. (1991). Stroke rehabilitation: A model predicting return home. *Western Journal of Medicine*, *154*, 587-590.

Whitlock, J. A., & Hamilton, B. B. (1995). Functional outcome after rehabilitation for severe traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, *76*, 1103-1112.

Yarushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Rep*, *62*, 1432-1449.

Appendix A

BNI SCREEN FOR HIGHER CEREBRAL FUNCTIONS
George P. Prigatano, Ph.D.

Patient Information

1. Patient Name:______________________________

2. DX:______________ 3. Age:__________ 4. D.O.B.:______________

5. Ed.______________ 6. Sex:__________ 7. Handedness:______________

8. Occupation:______________ 9. Estimated Premorbid IQ (Circle One - B, LA, A, HA, S

10. DOI/or onset of neurological problem:______________

11. Date of Exam:______________ 12. Chronicity:______________

13. Ad. GCS (if approp.):__________ 14. Hosp.#:______________ 15. NP #:______________

Pre-Screening Item	Score
Level of consciousness/alertness	3,2,1
Basic Language	3,2,1
Level of cooperation	3,2,1

Code	Screening Item	Score
A.	Fluent	1, 0
A.	Paraphasic	1, 0
A.	Dysarthric	1, 0
A.	Comprehension	2, 1, 0
A.	Naming	1, 0
D.	Visual Object Recognition	1, 0
A.	Repetition	2, 1, 0
A.	Reading	1, .5, 0
A.	Writing - Copying	1, 0
	- Dictation	1, 0
A.	Spelling - Irregular	1, 0
	- Phonetic	1, 0
B.	Right-left Orientation	1, 0
B.	Place	1, 0
B.	Date	1, 0
D.	Constructional Praxis	
	Dominant Hand	1, 0
	Nondominant Hand	1, 0
A.	Arithmetic - Alexia	1, 0
A.	- Dyscalculia	1, 0
C.	- Memory & Concentration	1, 0
C.	Digits - Forward	1, 0
	- Reverse	1, 0
D.	Visual Scanning	2, 1, 0
D.	Visual Sequencing	1, 0
D.	Pattern Copying	1, 0
D.	Pattern Recognition	1, 0
E.	Learning & memory: #-Symbol	4, 3, 2, 1, 0
F.	Affect Expression: (Angry/happy)	1, 0
F.	Perception of facial affect	1, 0
F.	Affect control	1, 0
F.	Spontaneous Affect	1, 0
E.	Delayed Recall - House	1, 0
	- Tree	1, 0
	- Flying	1, 0
G.	Awareness vs. Performance	1, 0

SNIS Total Score = ______/50

Code	BNI Screen Subscale Item	Score
A.	Speech & Language Functions	______ (Max. = 15)
B.	Orientation	______ (Max. = 3)
C.	Attention/Concentration	______ (Max. = 3)
D.	Visuospatial & Visual Problem Solving	______ (Max. = 8)
E.	Memory	______ (Max. = 7)
F.	Affect	______ (Max. = 4)
G.	Awareness vs. Performance	______ (Max. = 1)
	Total Subscale Score (Total Max. Subscale Score)	______ 41

Comments/Behavioral Observations

Appendix B

GEORGE P. PRIGATANO, KIRAN AMIN, AND LESLIE D. ROSENSTEIN

Appendix
BNI Screen for Higher Cerebral Functions

Patient Information

1. Patient Name: ______
2. DX: ______ 3. Age: ______ 4. D.O.B.: ______
5. Ed. ______ 6. Sex: ______ 7. Handedness: ______
8. Occupation: ______ 9. Estimated Premorbid IQ (Circle One - B, LA, A, HA, S)
10. DOI/or onset of neurological problem: ______
11. Date of Exam: ______ 12. Chronicity: ______
13. Ad. GCS (if approp.): ______ 14. Hosp.#: ______ 15. NP #: ______ 16. Examiner ______

	Pre-Screening Item	Score
	Level of consciousness/alertness	3,2,1
	Basic Language	3,2,1
	Level of cooperation	3,2,1

Code	Screening Item	Score
A.	Fluent	1, 0
A.	Paraphasic	1, 0
A.	Dysarthric'	1, 0
A.	Comprehension	2, 1, 0
A.	Naming	1, 0
D.	Visual Object Recognition	1, 0
A.	Repetition	2, 1, 0
A.	Reading	1, .5, 0
A.	Writing	
	Copying	1, 0
	Dictation	1, 0
A.	Spelling	
	Irregular	1, 0
	Phonetic	1, 0
B.	Right-left Orientation	1, 0
B.	Place	1, 0
B.	Date	1, 0
D.	Constructional Praxis	
	Dominant Hand	1, 0
	Nondominant Hand	1, 0
A.	Arithmetic	
	Alexia	1, 0
A.	Dyscalculia	1, 0
C.	Memory & Concentration	1, 0
C.	Digits	
	Forward	1, 0
	Reverse	1, 0
D.	Visual Scanning	2, 1, 0
D.	Visual Sequencing	1, 0
D.	Pattern Copying	1, 0
D.	Pattern Recognition	1, 0
E.	Learning & memory: #-Symbol	4, 3, 2, 1, 0
F.	Affect Expression: (Angry/happy)	1, 0
F.	Perception of facial affect	1, 0
F.	Affect Control	1, 0
F.	Spontaneous Affect	1, 0
E.	Delayed Recall	
	House	1, 0
	Tree	1, 0
	Flying	1, 0
G.	Awareness vs. Performance	1, 0

BNIS Total Score = ______/50

Code	BNI Screen Subscale Item	Score
A.	Speech & Language Functions	______ (Max. = 15)
B.	Orientation	______ (Max. = 3)
C.	Attention/Concentration	______ (Max. = 3)
D.	Visuospatial & Visual Problem Solving	______ (Max. = 8)
E.	Memory	______ (Max. = 7)
F.	Affect	______ (Max. = 4)
G.	Awareness vs. Performance	______ (Max. = 1)
	Total Subscale Score	______
	(Total Max. Subscale Score)	41

Comments/Behavioral Observations

Appendix C

A total maximum score of 50 points is possible when scoring the BNIS (Prigatano, Amin, & Rosenstein, 1993a). The Speech and Language Functions subscale is composed of 13 items producing a maximum score of 15 points. Basic language abilities, such as speech comprehension, object naming, and sentence repetition, as well as secondary language functions, such as reading, spelling, writing, and written arithmetic are assessed. The Orientation subscale is comprised of 3 items resulting in a maximum score of 3 points, with orientation to place, date, and right-left body space assessed. The Attention/Concentration subscale has 3 items with a maximum score of 3 points. This is composed of five digits forward and five digits backward repetition tasks and a mental arithmetic task. The Visuospatial and Visual Problem Solving subscale is made of 7 items resulting in a maximum score of 8 points. This subscale assesses object recognition, visual scanning, visual sequencing, pattern copying and recognition, and constructional praxis. The Memory subscale is composed of 2 items, resulting in a maximum score of 7 points. The first item is a 4 point visual learning and recall task, with the second item being the recall of 3 words after time delay and distracter tasks. Affect is assessed through the administration of 4 items, producing a 4 point maximum score. The ability to control and express spontaneous affect is scored, along with the ability to correctly perceive and verbally label line drawings of facial affect. Finally, Awareness is rated on a one point scale, with the discrepancy between the subject's actual recall of the three words being compared to their pre-determined estimation of recall. During the initial presentation of the 3 words to be recalled at a later time the subject is subsequently asked to guess as to how many words they will actually recall. Any discrepancy, either too high or too low, results in a zero score for the Awareness subscale. The total subscale scores result in a maximum of 41 points. These are then added to the three prescreening item scores (maximum of 3 points each) for a total possible maximum score of 50 points.

Appendix D

Index measure

		Subject has condition	Subject doesn't have condition	
Test results from the measure in question	Positive: subject has the condition	a)True Positive	b) False Positive	a + b
	Negative: subject doesn't have the condition	c) False Negative	d)True Negative	c + d
		a + c	b + d	

Sensitivity = a / (a + c)

Specificity = d / (b + d)

False Positive rate = b / (b + d)

False Negative rate = c / (a + c)

Positive Predictive Power = a / (a + b)

Negative Predictive Power = d / (c + d)

Appendix E

FIM + FAM Items

- Motor Items
  - Self Care
    - Eating
    - Grooming
    - Bathing
    - Dressing upper body
    - Dressing lower body
    - Toileting
    - Swallowing *
  - Sphincter Control
    - Bladder Management
    - Bowel Management
  - Mobility
    - Bed, chair, wheelchair transfer
    - Toilet transfer
    - Tub, shower transfer
    - Car transfer *
    - Walking, wheelchair locomotion
    - Stairs
    - Community mobility *
- Cognition Items
  - Communication
    - Comprehension
    - Expression
    - Reading *
    - Writing *
    - Speech intelligibility *
  - Psychosocial Adjustment
    - Social interaction
    - Emotional status *
    - Adjustment to limitations *
    - Employability
  - Cognitive Function
    - Problem solving
    - Memory
    - Orientation *
    - Attention *
    - Safety judgement *

* FAM items

Appendix F

FIM + FAM 7 Point Scale

INDEPENDENT	Another person is not required for the activity (No Helper)
7	Complete Independence: All of the tasks described as making up the activity are typically performed safely, without modification, assistive devices, or aids, and within a reasonable amount of time.
6	Modified Independence: One or more of the following may be true: the activity requires an assistive device; the activity takes more than reasonable time, or there are safety (risk) considerations.
DEPENDENT	Subject requires another person for either supervision or physical assistance in order for the activity to be performed, or it is not performed (requires helper).
Modified Dependence	The subject expends half (50%) or more of the effort. The levels of assistance required are:
5	Supervision or Setup: Subject requires no more help than standby, cuing or coaxing, without physical contact or , helper sets up needed items or applies orthoses.
4	Minimal Contact Assistance: Subject requires no more help than touching, and expends 75 % or more of the effort.
3	Moderate Assistance: Subject requires more help than touching, or expends half (50%) or more (up to 75%) of the effort.
Complete Dependence	The subject expends less than half (50%) of the effort. Maximal or total assistance is required, or the activity is not performed. The levels of assistance required are:
2	Maximal Assistance: Subject expends less than 50 % of the effort, but at least 25%.
1	Total Assistance: Subject expends less than 25 % of the effort.

(from Guide for the Uniform Data Set for Medical Rehabilitation, 1993).

Table 1.

BNIS Predictive Power Results

Article	Subjects	Sensitivity	Specificity	PPP	NPP
BNIS Manual, 1993 (Total score)	BNI patients (initial validation) N = 188	89 %	50 %	82 %	63 %
(same as above)	Delayed Recall only	65 %	90 %	95 %	49 %
	Awareness only	71 %	92 %	96 %	55 %
BNIS Manual, 1993 (Total score)	BNI patients (expanded validation) N = 355	93 %	56 %	84 %	76 %
(same as above)	Delayed Recall only	68 %	93 %	96 %	54 %
	Awareness only	69 %	91 %	95 %	54 %

Table 2.

Inter-rater Reliability of the Functional Assessment Measure Items

Item Name	Domain	ICC
11. TRANSFER- TOILET *	Transfers	0.97
10. TRANSFER- BED, CHAIR, W/C *	Transfers	0.97
15. STAIRS *	Locomotion	0.97
12.TRANSFER- TUB / SHOWER *	Transfers	0.95
7. SWALLOWING	Self-Care	0.95
4. DRESSING (UPPER) *	Self-Care	0.95
1. FEEDING *	Self-Care	0.93
5. DRESSING (LOWER) *	Self-Care	0.93
14. WALKING / WHEELCHAIR *	Locomotion	0.92
13. TRANSFER - CAR	Transfers	0.92
6. TOILETING *	Self-Care	0.91
2. GROOMING *	Self-Care	0.89
9. BOWEL MANAGEMENT *	Sphincter Control	0.87
8. BLADDER MANAGEMENT *	Sphincter Control	0.84
3. BATHING *	Self-Care	0.84
30. SAFETY JUDGEMENT	Cognitive Functioning	0.82
28. ORIENTATION	Cognitive Functioning	0.82
19. READING	Communication	0.82
18. EXPRESSION *	Communication	0.81
25. EMPLOYABILITY	Psychosocial Adjustment	0.80
21. SPEECH INTELLIGIBILITY	Communication	0.79
20. WRITING	Communication	0.79
26. PROBLEM SOLVING *	Cognitive Functioning	0.78
27. MEMORY *	Cognitive Functioning	0.78
16. COMMUNITY MOBILITY	Locomotion	0.75
24. ADJUSTMENT TO LIMITATIONS	Psychosocial Adjustment	0.74
17. COMPREHENSION *	Communication	0.70
29. ATTENTION	Cognitive Functioning	0.68
23. EMOTIONAL STATUS	Psychosocial Adjustment	0.62
22. SOCIAL INTERACTION *	Psychosocial Adjustment	0.36

* FIM Items

VITA AUCTORIS

Peter Wass was born in Scarborough, Ontario on August 24, 1959. He obtained an Honours B.A. in Psychology in 1989 from Laurentian University. From there he went on to the University of Windsor in the Clinical Neuropsychology Graduate School Program. He obtained his Master's Degree from the University of Windsor in 1991. He is currently a candidate for the Ph.D. degree in Clinical Neuropsychology at the University of Windsor and hopes to graduate in the Fall of 1997.