NEUROPSYCHOLOGICAL ASSESSMENT OF COGNITIVE FUNCTIONING IN INDIVIDUALS WITH EXPRESSIVE DISABILITIES IN ADDITION TO TRAUMATIC BRAIN INJURIES

A dissertation presented in partial fulfilment of
the requirements of the degree of Doctor of Philosophy
in Psychology at Massey University.

Duncan Ross Babbage
2000

He will wipe every tear from their eyes.
There will be no more death or mourning or crying or pain,
for the old order of things has passed away.

— The Bible, Revelation 21:4

ABSTRACT

This research programme focussed on individuals who have severe physical and sensory disabilities that interfere with standard neuropsychological assessment. Current assessment guidelines when working with people who have such disabilities were examined, which revealed that while various suggestions have been made, much work is required to empirically evaluate the most appropriate procedures for conducting such assessments. The current research was an attempt to further examine these issues.

The first study, a retrospective review of a cohort of individuals referred for neuropsychological assessment after traumatic brain injury, was conducted to determine the focus population for the research. Based on this review, the research was limited to individuals who (in addition to traumatic brain injuries) had severe *expressive* disabilities. That meant they were unable to speak, write, draw, or manipulate test materials—the common modalities for making responses in neuropsychological assessment of cognitive functioning. This review also examined the types of adaptations to standard measures that were required in order to assess such individuals.

The research questions related to whether a comprehensive assessment across the domains of cognitive functioning could be undertaken with people in this group. Therefore, *comprehensive* cognitive assessment had to be defined. The next study therefore examined the issue of what constitutes a comprehensive neuropsychological assessment of cognitive functioning, by evaluating the domains into which researchers divided cognition in their journal articles. All articles published in four neuropsychology journals over a 12-month period were reviewed. Based on this information, a formulation of the domains of cognitive functioning was developed.

In the third study was a survey of neuropsychological practitioners, in which a case vignette of an individual which severe expressive disabilities was presented. Respondents were asked to discuss the assessment strategies they would use in such a case. In addition, the survey examined whether clinicians divided the assessment of cognitive functioning into the same domains identified in the earlier journal review. The case vignette discussions provided suggestions regarding assessment strategies for people with expressive disabilities. The survey provided support for the earlier formulation of the domains of cognitive functioning developed from the research review. This formulation was used, therefore, in selecting the domains to be assessed in the final clinical phase of the research.

Based on these studies, a group of measures was selected and adapted that would be suitable for the assessment of cognitive functioning in individuals with expressive disabilities. These measures covered a broad range of cognitive domains allowing for as comprehensive an assessment as possible, while not requiring an examinee to speak, write, draw, or manipulate test materials. Rather, an examinee was required only to select from multiple-choice answers by pointing, or in some cases to spell out answers on an alphabet board.

These measures were administered to three groups of participants: individuals with expressive disabilities in addition to traumatic brain injuries, individuals with traumatic brain injury alone, and a healthy normative group. To examine the psychometric properties of these measures, a group of comparison measures, administered in their standard formats, was also included in the protocol. These tests were selected to measure the same constructs as the adapted measures, and were used to provide a benchmark against which performance on the new measures could be evaluated. As these comparison measures were administered in standard format, they could only be administered to the non-disabled participants. To evaluate further the adapted measures, some participants were seen for follow-up assessments two to four weeks after their initial assessments, and the adapted measures were re-administered.

Internal consistency and test-retest reliability of the measures were investigated, and concurrent, construct and discriminant validity were also examined. The measures in this protocol were generally found to be reliable and valid neuropsychological assessment instruments and the results provided support for the types of adaptations trialed in this research. The performances of individuals with expressive disabilities were examined closely, which indicated that people with these disabilities were able to manage the task requirements of the adapted tests and that the tests were generally of appropriate difficulty. Qualitative aspects of conducting assessments with people with expressive disabilities were also discussed.

Based on the performance of individuals in the normative group, preliminary norms were presented as both standard scores and percentile scores. These data were presented so that clinicians using the adapted measures described in this research could compare examinees to a reference group. However, the adaptations could not be considered fully standardised measures and the limitations of both the tests and the norms were discussed.

The final section reviewed the aims that were outlined at the beginning of this research programme. Each of the 20 objectives of the research were met. The research provided clinically relevant information about working with people with severe expressive disabilities, was conducted in an ethical manner, which considered carefully the specific needs of participants, particularly those with severe disabilities, and did so in a manner that maintained scientific rigour and objectivity.

ACKNOWLEDGEMENTS

Firstly, I acknowledge the participants. They gave substantial time and effort for little personal return. While many reported enjoying the assessments, participants seemed widely to hold one motivation: to help improve things for people with severe brain injuries in the future. The level of interest in this project was both encouraging and humbling. Without the support of the staff at each rehabilitation service that I worked through, this project would have been considerably more difficult.

I thank the Health Research Council of New Zealand, who supported the clinical phase of this research with a Clinical Research Limited Budget Grant. A project of this size would not have been possible without this funding. Thank you also to the New Zealand Vice-Chancellors' Committee for their support of me personally with a William Georgetti Scholarship and a Shirtcliffe Fellowship, and Massey University for the award of a Massey Doctoral Scholarship.

Thanks to Angela Person, Angela McNaught, and Matthew Manderson, who were employed as research assistants on this project. It was great to be working with such capable people. Thank you for your hard work.

Professor Janet Leathem was first supervisor of this project. Her considerable clinical and research experience were essential, and her relentless positive attitude was a great support. Thanks, Janet. It has been excellent working with you.

Dr. John Podd was second supervisor of this research and provided outstanding advice, practical, statistical and experimental. I valued your help a great deal John, and you often gave me cause to think. Thank you.

Thanks to Associate Professor John Spicer, who was consulted on a number of occasions from initial design through to final analysis. John, your ability and willingness to communicate your broad statistical knowledge was greatly appreciated.

Finally, I wish to acknowledge the support of my wife, Bronwyn, who never once questioned my need to spend the remainder of the millennium conducting this research. She tolerated the inevitable highs and lows, the long periods of time I spent away from home during the clinical phase data collection, and the unbalanced lifestyle that went with completing the writing. I always felt her support.

I would also like to note that during the final write-up stage of this research I have begun considering a new hypothesis (unrelated to the current project). After the success of many long writing sessions into the small hours of the morning, I have postulated that procrastination may be solar powered. A focus for future research?

May 2000

TABLE OF CONTENTS

PART III: CLINICAL PHASE: EXAMINATION OF ASSESSMENT FRAMEWORK 107

LIST OF TABLES

LIST OF FIGURES

PART I:

FOUNDATIONS, AIMS AND OBJECTIVES

PREFACE

This research examines the issues associated with the assessment of cognitive functioning in people who, in addition to neurological insult, have severe physical and sensory disabilities. The research was driven by a number of initial aims. The first was a desire to conduct a piece of research that was clinically focussed, and involved working with a clinical population. The intention was further to produce a piece of work of relevance to the individuals who participated in the project; if not of benefit to them personally, then to those with similar disabilities in the future. Finally, it was the intention in conducting the research to make a contribution, though of necessity small, to the knowledge of the field of clinical neuropsychology.

The dissertation is divided into four parts. Part I describes the basis for the research, outlining in detail the original rationale, and the specific aims and objectives that were set.

Part II reports three studies that were conducted to narrow the focus of the research programme. These three studies identify the specific participant population, the domains of assessment to be focussed upon, and the manner in which assessment instruments should be adapted. On the basis of these studies, that section concludes with the selection of the research measures to be included in the clinical phase.

The clinical phase of the research, Part III, was a study that examined the psychometric properties of a group of cognitive assessment measures, adapted for use with people with severe disabilities. The properties of these measures were examined in a group of individuals with severe traumatic brain injuries and in another with traumatic brain injuries in addition to severe disabilities. As part of this study, a substantial normative group was also sampled. This study was conducted in order to examine the utility of these measures for future neuropsychological assessment.

In the final section of the dissertation, the research is evaluated to determine whether the objectives defined at the outset were fulfilled, and thus the extent to which the research achieved its aims.

CHAPTER 1: FOUNDATIONS OF THE RESEARCH PROGRAMME

Introduction

This chapter lays the foundations for the current research programme—an examination of the neuropsychological assessment of individuals who, as well as experiencing neurological insults, also have severe physical and sensory disabilities that interfere with the standard assessment approach. The purpose of neuropsychological assessment is first discussed, along with factors that at times limit its usefulness. Recognising the diversity of the field, the philosophical basis of this research is then described. The assessment of people with physical and sensory disabilities is examined, and current practices are compared to the ideal. In this context, the current research programme is outlined. The chapter concludes by laying out the research questions that the programme endeavours to answer.

Purpose of Neuropsychological Assessment

Neuropsychological assessment is conducted for many purposes, for example to investigate the presence of an organic condition, localise a lesion, assess work capacity, plan education and rehabilitation, establish baselines, aid in discharge planning, and to assist with forensic questions (Guilmette, Faust, Hart, & Arkes, 1990). Aside from this and in a more general sense, assessments are also conducted to determine the extent of functional disability resulting from an insult, establish whether an individual is capable of living independently and of managing their legal and financial affairs, to predict likely prognosis following an insult, and for research purposes.

A core part of the neuropsychological assessment process is the administration of cognitive neuropsychological testing. Other parts of the process typically also include assessment of personal and interpersonal functioning, both through interview, and by the administration of psychometric assessment instruments such as personality measures (e.g., Minnesota Multiphasic Personality Inventory-II; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) or measures of mood or anxiety (e.g., Beck Depression Inventory-II, Beck, Sterer, & Brown, 1996; State-Trait Anxiety Inventory, Speilberger, 1983). The ability of psychologists to provide information based on tests which have demonstrated reliability and validity represents a unique and important contribution to the care of individuals who have experienced neuropathological insults (Prigatano & Redner, 1993).

Limitations of Neuropsychological Assessment

However, neuropsychological assessment is not without limitations. Firstly, conclusions arising from assessment are necessarily limited by our current understanding of brain-behaviour relationships (Levin, 1994). Challenges to the reliability and validity of particular tests, and of assessment practices in general, continue to be raised (e.g., Faust, 1991; Heinrichs, 1990), including particularly issues associated with clinical judgement and the clinical decision making process (e.g., Faust et al., 1988; c.f. Garb & Schramke, 1996; Kleinmuntz, 1990). Contextual limitations are important determinants of clinical assessment practice. Limited financial and personnel resources affect the ability to translate theoretically ideal assessment into clinical practice. Additionally, progress establishing the ecological validity of neuropsychological assessment has been slow, although there have been efforts in this direction (e.g., Sbordone & Long, 1996). Similarly, recognising that individual variables such as age, cultural background and socio-economic status may affect assessment results (Ardila, 1995; Binder & Thompson, 1995; Segall, 1986), significant efforts are being made to examine practice (e.g., Puente, 1990; Shepherd & Leathem, 1999). This has included adapting or developing tests that are less sensitive to these confounding factors (such as changes in the third edition of the Wechsler scales to remove some items specific to North American culture; Wechsler, 1997a), and conducting normative studies which recognise the diversity of the populations with which the tests may be used (e.g., Ivnik, Malec, Smith, Tangalos, & Petersen, 1996; Ivnik et al., 1997).

In addition to the factors discussed above, elements of the client's medical and personal history may impact on their ability to complete neuropsychological assessment. Neuropsychological tests have generally been constructed for use with physically able individuals (Lezak, 1995) and are not as easily applied when assessment is required of individuals with physical or sensory disabilities, where standard methods of stimulus presentation and response modality may be incompatible (Levin, 1994).

Provision of assessment services to individuals with physical and sensory disabilities is extremely important. Many of these individuals may be cognitively intact, or at least have areas of intact functioning. Clarification of areas of residual strength and weakness can direct rehabilitative efforts (Gass & Brown, 1992) to provide the highest quality of life that is possible. Additionally, where neuropathology has been established, assessments can establish the degree of control an individual should have over their financial and legal affairs, and to determine whether they are capable of living partly or completely independently.

Various approaches have been taken to the assessment of cognitive functioning in individuals with severe disabilities. For example, Goldstein (1987) supported the use of fixed neuropsychological batteries in such assessments, but also argued for the systematic and regular use of behavioural assessment strategies in neuropsychological rehabilitation. Other authors have argued for a move away from rigid adherence to psychometric traditions in favour of more flexible approaches to psychometric assessment (e.g., Kiernan, 1989), or have called for a greater reliance upon subjective clinical judgements, viewing the assessor as "an active constructor and evaluator of hypotheses rather than merely an administrator of pre-packaged tests" (Schuler & Perez, 1991, p. 105). However, as stated earlier, while clinical judgement can be brought to bear on such issues, the view taken for the purposes of this research is that where available, standardised assessment can provide unique information not available from other sources.

In the clinical setting where this research was undertaken the philosophical approach to the selection of neuropsychological assessment measures is a flexible battery approach. This approach involves the use of a routine group of tests for particular types of patients, such as people with traumatic brain injuries, or stroke, after which further clarification of an area might be undertaken using additional testing. In recent years this approach has gained increasing popularity, and in a large 1999 survey of neuropsychologists in the United States 69% of respondents endorsed this approach (J. Sweet, personal communication, 20 August 1999; see Chapter 5 for further discussion of this issue). However, in a research context a flexible battery approach is less suitable, and in the clinical phase of this research programme a fixed research protocol was employed.

The current research is based in traditional psychometric theory of psychological and neuropsychological assessment. While alternative models of assessment have at various times been proposed (e.g., testing based on Bayesian theory; Elwood, 1993; Jones, 1989) current practice in the field is still firmly grounded in traditional psychometric theory.

Assessment Of Individuals With Severe Disabilities

Idealised Assessment

In 1985 the American Psychological Association, in conjunction with the American Educational Research Association and the National Council on Measurement in Education, released benchmark standards for the use of educational and psychological testing (American Psychological Association, 1985). In particular, these standards included a section on the assessment of people who have disabilities. Specific issues for consideration in testing such people included that:

- Modification of tests is often required and usually desirable. However, it cannot be assumed that the modified measures have the same psychometric properties and reliability and validity should be established for such modified measures. It cannot be assumed that modified measures are even tapping the same construct.
- In some cases, individuals with disabilities may have had a different developmental history due to their disability. Therefore, if a test examines information primarily learned through vision it may be inappropriate for people with visual impairment, even if it is modified so that they are able to complete the basic task requirements.
- Ideally neuropsychological assessment of individuals with physical and sensory disabilities would be based on measures which have been adapted specifically to be suitable for people with that disability (e.g., visual impairment, particular physical disabilities), have been demonstrated to be reliable and valid, and for which appropriate normative information has been collected.

- Normative information for such tests might be based on other people with similar disabilities, or the general population, depending on the purpose of the assessment (American Psychological Association, 1985).

Current Assessment Practice

Neuropsychologists commonly assess people who have various forms of disability related (or in some cases unrelated) to their neuropathology. While the reasons for conducting assessment vary, a primary function of neuropsychological assessment is usually the assessment of cognitive disability as a result of such insults to the brain. Therefore, it would seem logical to assume that in comparison to other disciplines where testing is predominantly conducted with the general population (e.g., in education), neuropsychologists would be well prepared to work with people with physical and sensory disabilities. Two of the issues that must be considered in such work are the changes in cognitive development that may be due to disabilities (Mayberry, 1992; Warren, 1992), and specific changes that must be made to the process of assessment. However, while Levin (1994) warns that such disability factors unrelated to a person's cognitive functioning may produce spuriously low results on some tests, most clinicians have not received specific training in working with people with this level of disability (Golden & Robbins, 1985), and have little experience in such work (Swiercinsky, 1987).

In the absence of specific training and established guidelines, assessment of people with physical and sensory disabilities usually involves the cautious (and sometimes ingenious) administration of such traditional neuropsychological assessment measures as it is possible to administer without compromising the original psychometric properties (Lezak, 1995). When traditional tests are used with individuals with disabilities, care must be taken in making comparisons between a person's performance and non-disabled normative groups. While these comparisons are not automatically invalid, attention must be paid to the way normative information is used (Swiercinsky, 1987). For example, if an individual with a severe motor impairment performs at low levels on a timed measure such as Block Design (WAIS-III; Wechsler, 1997a), the performance may be compromised by motor rather than cognitive factors (see Berent et al., 1990; Berent et al., 1992; Lalonde, Botez, &

Botez, 1992; Reitan & Wolfson, 1994a) and it may not be possible to separate these effects. However, where the focus is not on cognitive abilities, but rather on estimating function in a working environment compared to non-disabled colleagues, use of standard normative information may be appropriate (American Psychological Association, 1985; Swiercinsky, 1987; Wilson, 1987).

Despite additional complications in the assessment process, it should still be possible to conduct a comprehensive neuropsychological assessment with individuals with physical and sensory disabilities. A number of authors have outlined the issues associated with the assessment of a range of sensory and motor deficits, providing some initial starting points for clinicians confronted with a referral to assess a person with such disabilities (e.g., Golden & Robbins, 1985; Lezak, 1995; Swiercinsky, 1987; see also Botterbusch, 1976; and Mittler, 1970, for early reviews of the topic). These sources discuss specific alternative tests that have been published for some areas (e.g., the Haptic Intelligence Scale for the Adult Blind; Jordan, 1978), and adaptations to measures that may be made by clinicians in other cases.

Guidelines for Assessment

In one of the more detailed treatments to date, Swiercinsky (1987) attempted to "offer both general guidelines and appropriate test procedures or modifications to use" (p. 198) when examining individuals with physical or sensory disabilities. He detailed five principles that should be considered in the selection and interpretation of neuropsychological assessment instruments with such populations:

- *Caution in the use of normative data*, and the interpretations drawn from this, with a clear understanding of the effects of disabilities on test performance.
- *Native communication* mechanisms should be used (such as sign language for a signing deaf person). Task requirements should be commensurate with the response modalities available to the person being tested, and areas examined should be domains that an examinee is familiar with.
- *Socialisation background* of individuals with disabilities may be different to that of the general population, and may result in limited knowledge in some areas. This should be taken into account.

- *Time since impairment* should allow sufficient time for an individual to adjust to the significant changes they have experienced.
- *Relation of impairment to brain damage* must also be considered, with onset of brain injury at the time of the physical or sensory disability leading to different coping strategies and psychosocial outcomes than someone who has experienced the disability for some time, and only recently acquired a brain injury.

Adaptations to Assessment

A number of other authors have called for clinicians to adapt assessment methods when examining people with disabilities, and have suggested possible adaptations (e.g., Geisinger & Carlson, 1995; Golden & Robbins, 1985; Kiernan, 1989). The history of such adaptations to accommodate the specific testing requirements of people with disabilities was reviewed by Caplan and Shechter (1995). They argue that modifications to standard assessment procedures are not just desirable, but necessary for valid, fair and ethical assessments, which follow principles enshrined in (American) law (see, for example, Ebner, Burkhead, & Merydith, 1994).

However, while the suggestions made by many authors based on considerable clinical experience in the area are helpful, the literature appears to contain more articles calling for clinicians to adapt their assessment with these populations, and suggesting possible adaptations, than studies which empirically scrutinise such adaptations. In a review of the 1992 revision of the Ethics Code of the American Psychological Association and its implications for neuropsychological assessment, Binder and Thompson (1995) recommended that such adapted procedures should only be used in the absence of standardised alternatives, and that examiners should proceed “with caution” (p. 32), interpreting the results conservatively.

Even the WAIS-R as a Neuropsychological Instrument (WAIS-R-NI; Kaplan, Fein, Morris, & Delis, 1991), adapted from the Wechsler Adult Intelligence Scale, was published without any studies reporting its use in patient populations. In 1995 Caplan and Shechter wrote that they knew of no studies published in any patient population using this scale, but argued that due to the nature of the scale (focussed on flexible adjustment of the WAIS to meet individual needs) a systematic psychometric standardisation would not be possible. However, they called for empirical examination of the individual adaptations that may be made using the measure. At the time of writing only three studies could be located which had conducted any research using the measure (Connolly, Mate-Kole, & Joyce, 1999; Slick et al., 1996; Troester, Fields, Paolo, & Koller, 1996), one a single case study (Connolly et al.), the others both with small numbers of participants and only one of these (Troester et al.) using a clinical population.

When studies have reported evaluations of other adapted assessment procedures, these are often confined to single tests or areas of functioning. For example, Caplan and Shechter (1995) describe three such studies they conducted, and similar studies by other authors; other examples can be found in Beaumont, Marjoribanks, Flury, and Lintern (1999) and in Smith, Cash, Barr, and Putney (1986). While confining the focus of research articles to single tests or domains is not uncommon, this adds to the challenge facing a clinician who might attempt to develop, on the basis of such studies, a comprehensive assessment strategy for use with clients with a particular group of disabilities.

More broad-based cognitive functioning has been examined in some studies of individuals with spinal cord injuries who may have sustained brain injuries (e.g., Richards, Brown, Hagglund, Bua, & Reeder, 1988; Roth et al., 1989; Wilmot, Cope, Hall, & Acker, 1985). These studies report a 'comprehensive' battery of tests, which have been selected on the basis of only requiring a verbal response.

One area where there has been significant research over some period of time is in the cognitive assessment of children with cerebral palsy (see, for example, Blum, Burgemeister, & Lorge, 1951; Sattler & Tozier, 1970; Schonell, 1958; for three early reviews; also Anastasi, 1988; Coop, Eckel, & Stuck, 1975; Dorman, Hurley, & Laatsch, 1984; Dorman, Laatsch, & Hurley, 1985). Typically, such studies appropriately use measures that were designed for children, unfortunately making most of them unsuitable in the assessment of adults with this level of disability. In addition, as individuals with cerebral palsy have an altered developmental history from birth, their cognitive organisation may be considerably different to adult individuals who acquire (grossly) similar physical and sensory disabilities as adults. Consequently, extrapolation from these studies is confounded. Nonetheless, while cognitive assessment measures have changed considerably in the past 50 years since the early reviews mentioned above, the recommended adaptations to testing procedures vary little from those offered in more contemporary literature (for example, compare Botterbusch, 1976; Heaton & Heaton, 1981; Sattler & Tozier, 1970; Schonell, 1958; Swiercinsky, 1987; and the recommendations in Lezak, 1995). Furthermore, over this considerable period, little research has systematically examined the effects of such adaptations on test validity.

Research on Multiple Disabilities

As well as a paucity of research examining such adaptations, little research has been conducted in general on the neuropsychological assessment of adults with multiple severe physical and sensory disabilities. One reason for the lack of a well-established body of empirical research in this area may be that very few people survive with this level of disability. Even those working with clients with severe neuropathology would be unlikely to see people with multiple severe physical and sensory disabilities in sufficient numbers to develop a depth of experience working with this group. Thus, while ad hoc adaptations to instruments may provide valuable information regarding a client's cognitive functioning, it is unlikely that such adaptations could ever be systematised through this route.

Initial forays have been made into research on the assessment of adults with severe physical and sensory disabilities. Some researchers, such as Kreutzer, Gordon, Rosenthal, and Marwitz (1993) have now begun to include individuals with sensory and motor disabilities in their research samples by using measures which do not rely on visual and motor functioning. Magner and Spector (1984) reported a case study of a profoundly response-limited adult, for whom a range of measures were adapted. These included tests of memory, verbal functions, non-verbal reasoning, concept formation and abstraction, visuoperception, and academic achievement. All the measures were multiple-choice in their standard administration, and were adapted to be used with a yes/no response. An assistant, where possible blind to the correct answers, pointed to each response in turn, and the examinee responded either yes or no, using a system of grunts. The adapted procedures included a defined period in which the examinee had to respond for answers to be recorded. Using these procedures the authors reported the assessment provided information useful in modifying the individual's treatment plan, and noted that without such adaptations the examinee would have been untestable.

Berninger, Gans, St. James, and Connors (1988) conducted perhaps one of the best studies in this area, trialing an adapted version of the WAIS-R (Wechsler, 1981) for use with people with speech or upper motor disabilities, or both. Their first study examined the performance of 16 individuals with severe physical disabilities on both adapted WAIS-R subtests and the parts of the standard WAIS-R they were able to complete. Performances of this group were compared to those of a healthy normative group (n=32). A second study compared 14 people with cerebral palsy (8 of whom had participated in the initial study) to a group of 14 people with severe spinal cord injury (see article for full details). Modifications involved the following adaptations:

- Adaptation to a multiple-choice format, where participants could respond verbally, by pointing, or simply with a yes/no response.
- Enlargement of stimuli. For Picture Completion and Picture Arrangement, this was a simple two-dimensional enlarged picture. Block Design and Object Assembly pieces were similarly enlarged, and presented on a magnetic board.
- For subtests that normally require manipulation by the examinee, participants answered yes/no questions from the examiner regarding manipulations, which were then made by the examiner.
- Timing was removed, and thus so were bonus points for speed.

The authors of these studies concluded that this modified WAIS-R showed promise "as a research tool" (Berninger et al., 1988, p. 254). Unfortunately, no further research on this adaptation was conducted (Caplan & Shechter, 1995).

Closing the Gap — Current Versus Ideal Neuropsychological Assessment

On the basis of this review of the literature, it would appear that current neuropsychological assessment practice for people with severe disabilities varies considerably from what has been identified as the ideal[1]. While many authors over a long period have made clinical suggestions for conducting neuropsychological assessment of people with disabilities, these have been subjected to scant empirical evaluation. Some research has been conducted on specific populations using specially constructed measures, which have limited generalisability. Understandably, the least research has been conducted with the most severely disabled populations, particularly individuals in which more than one sensory or expressive modality has been affected.

The current research represents an initial attempt to rectify this imbalance. In doing so it was considered important that the research be clinically relevant, focussing directly on the questions that a clinician might be asked when referred a person with severe physical and sensory disabilities.

The Current Project — A First Step

Initial Research Question

This research was undertaken to attempt to make some progress in closing the gap between the stated ideal assessment procedures (American Psychological Association, 1985) and those currently available to clinicians. The research aimed to answer the question, 'To what extent can a comprehensive neuropsychological assessment be conducted with individuals with physical and sensory disabilities, and how should this be accomplished?'

This research was conceived as being the first step, or perhaps another step, along the path of answering this question. However, such a research question is broad and untestable. In order to be able to seek answers, it was necessary to reduce the breadth of the question being asked, and to develop a series of more specific questions.

---

[1] It is recognised that some authors would argue that the application of standardised cognitive assessment to individuals with severe physical disabilities is far from ideal, and would advocate vastly different forms of assessment (e.g., behavioural observation). However, the term *ideal* in this context refers to the ideal manner in which to conduct an assessment, given that psychometric cognitive assessment methods are to be employed.

Sharpening the Focus

The current research was limited to an investigation of the use of *psychometric* testing for the examination of functioning. Many other forms of assessment (e.g., interviews, clinician rating scales, behavioural observation, use of collateral information, neuroimaging) certainly have their use with people with severe disabilities. Recent research, for example, is examining monitoring brainwave activity in response to stimulus presentation, through the measurement of event-related potentials, in order to conduct assessments of cognitive functioning where an individual does not have a reliable yes/no response (see Connolly, Mate-Kole, & Joyce, 1999). There is certainly a role for such technology, and no doubt procedures of this kind will become more common if they prove successful. Ultimately it seems likely that such assessments would be routinely extended to the use of functional neuroimaging, as it becomes increasingly accessible and sophisticated.

Psychometric assessment is not intended to answer all questions; other assessment methods do not compete with, but rather supplement, neuropsychological assessment. However, traditional psychometric assessment is more accessible to clinicians currently than technologies such as those described above, and is considerably less expensive. Therefore, an examination of the utility of psychometric assessments with clients with physical and sensory disabilities is deserving of further attention. The current study aims to investigate whether such assessments can be brought under the umbrella of standardisation.

In addition to concentrating on psychometric assessment, the focus of the research was further restricted to the psychometric assessment of *cognitive* functioning, as opposed to, for example, personal or interpersonal functioning. Assessment of these domains is an essential part of neuropsychological assessment (Lezak, 1995), and in no way should their exclusion in this project be taken to infer that personal and interpersonal functioning are considered less important for people with severe physical and sensory disabilities. However, there are methodological similarities in the administration of various cognitive assessment measures that are not shared with the assessment of personal and interpersonal functioning. In order to reduce the research to a manageable size, the decision was thus made to restrict the focus of the research to cognitive functioning.

Thirdly, the decision was made to limit the clinical phase of the research to include individuals with similar disabilities. In order to have a degree of group homogeneity, it was planned that individuals would only be included if they had experienced traumatic brain injuries, as opposed to other forms of neurological insult. Traumatic brain injury is a leading cause of neuropathology, and also a specialist area at the Psychology Clinic at Massey University. For both of these reasons, access to a significant client population was available by selecting this group.

The decision was also made to restrict the focus of the clinical phase of the research to a group of individuals who in a gross sense had relatively similar physical and sensory disabilities. Considering the small sample sizes available in research of this nature, and large individual variation, this restriction was not made in order to be able to look for 'group effects' *per se*, but rather to have a collection of individuals with whom similar kinds of adapted measures were likely to be suitable. The decision as to what the specific nature of these disabilities would be was not made *a priori*, but rather was the focus of the study reported in Chapter 3.

Final Research Questions

Having defined more closely the parameters of the population to be studied and the domains of investigation, the research question was similarly refined. The research aimed to answer the following questions:

In the selected group of individuals with severe physical and sensory disabilities,

- Are there any domains of cognitive functioning in which existing psychometric assessment measures are unsuitable?
- Can existing psychometric measures be adapted to be suitable, or new assessment measures created?
- Are there any domains for which no assessment is possible?
- What are the psychometric properties of the adapted and new tests, and do they meet the minimum criteria for acceptable measures?

- Are the psychometric properties of adapted measures different from the properties of the measures on which they were based?
- What information, if any, is lost though changes to measures subtly affecting the underlying construct they examine?

Additionally, two broader research questions were posed.

- What further research could productively be conducted in this area?
- What guidelines can be offered to clinicians who are conducting assessments with individuals with severe physical and sensory disabilities, particularly where the adaptations examined in the current research prove unsuitable?

Summary of Foundations of the Research Programme

The current review identified that previous literature has raised issues associated with the neuropsychological assessment of people with severe physical and sensory disabilities, but that much work is still required regarding the most appropriate procedures for conducting such assessments. The current research is an attempt to further examine these issues. The focus population for the research was clarified—a group of individuals with traumatic brain injuries, who shared similar physical and sensory disabilities. In addition, the research is limited to an examination of the assessment of cognitive functioning. The specific research questions to be answered were outlined.

CHAPTER 2: AIMS AND OBJECTIVES

Regarding Aims and Objectives

The previous chapter described the factors that led to the research being conducted, and the foundation on which it was built. A general question, which in its raw form was broad and untestable, was limited to a group of specific, testable, questions. In seeking to answer these, a series of aims were introduced which will be made explicit in this chapter. Two of the aims related to the product of the research efforts, while another two related to the manner in which it should be conducted. The former two can be seen as the reasons for which the research was conducted.

All of these aims were target ideals. Therefore, associated with each aim were also a series of objectives. These objectives stated the specific ways in which the research attempted to meet each of the aims. While the aims were ideals, the objectives were achievable and the extent to which the objectives were met could be evaluated.

It should be noted that these aims and objectives are in some senses arbitrary, and that other aims and objectives could have been selected. Already discussed is the decision to limit the research to the assessment of cognitive functioning. Similarly, the decision was made to restrict the focus of the clinical phase to individuals who in a general sense could be seen as a group with similar physical and sensory disabilities. This narrow focus was necessary to allow reasonable group comparisons to be made. However, it also means that the research may be less relevant to the assessment of individuals with other disabilities.

This dissertation will conclude with an evaluation of the extent to which the objectives were met, and thus the extent to which the aims were achieved.

Aims and Objectives of the Research Programme

Aim 1. That the research is clinically relevant, providing information about realistic assessment options for clinicians working with people with severe physical and sensory disabilities.

This means that the research should:

- Focus on a group of individuals who represent a significant proportion of the individuals with multiple disabilities who present to neuropsychologists for assessment.
- Have a broad-based focus that enables a comprehensive assessment to be made, rather than focussing solely on one domain of cognitive functioning.
- Use methods and equipment that practising clinical psychologists could access. Expensive or unwieldy custom equipment is therefore not suitable. Commercially published psychometric tests that are not commonly available would need to be carefully considered, since if a service sees clients with this level of disability only on rare occasions, they are unlikely to acquire such tests.
- Use timeframes that are realistic in clinical practice terms, regarding the length and number of assessment sessions required.
- Produce a group of adapted psychometric instruments that have been subjected to at least initial psychometric standardisation.
- Provide specific suggestions about adapting cognitive assessment measures, for use by clinicians where the measures developed are not suitable in particular cases.

Aim 2. That the research represents a significant contribution to knowledge regarding assessment of individuals with severe physical and sensory disabilities.

This means that the research should:

- Answer the research questions, as proposed in Chapter 1.
- Evaluate the extent to which the research questions were in fact the correct questions to ask in the light of the findings.
- Propose suitable areas for future research.
- Be disseminated in a way that informs psychologists, other health professionals, and consumers of health and disability services.

Aim 3. That the research conducted is ethical, does not cause harm to participants, and minimises any discomfort or distress they might experience either due to, or during, the assessment process.

This means that the research should:

- Be subjected to ethical review by peers and, where appropriate, ethical review committees.
- Utilise procedures for obtaining informed consent that are appropriate for the populations being studied, and obtain consent from as wide a group as possible. In the clinical groups, this might in some cases include guardians or caregivers. Under all circumstances, the informed consent would be obtained of participants themselves.
- Use assessment protocols that are as brief as possible while answering the research questions, and ensure that participants' needs are taken into consideration in the administration of the measures. Particular attention should be paid to the needs of participants with disabilities, such as in some cases the effects of increased fatigue and a more frequent requirement for personal care.
- Provide opportunity for participants to discontinue assessment at any time, and to not answer any particular question. Additionally, if the participant becomes upset by the assessment process, the examiner would discontinue assessment themselves, following discussion with the participant.

Aim 4. That the research is good science, maintaining rigour and objectivity.

This means that the research should:

- Operationally define all concepts in a manner consistent with both research and clinical practice.
- In the clinical phase, take reasonable steps to ensure that research measures are administered in a consistent way across individuals, groups, and examiners.
- Recognise that good science is also flexible, and that the best planned new assessment techniques may prove to be less than ideal when trialed with a clinical population. Nonetheless, the planned administration of measures will be adhered to, to the greatest extent possible. Any deviations from this should be carefully documented, and the implications of such deviations considered.

Value for Intervention and Rehabilitation

In addition to the aims above, it was hoped that any guidelines and techniques which were produced by the research would not simply inform assessment for its own sake, but would be valuable for guiding interventions with and rehabilitation of individuals with physical and sensory disabilities in the future. However, it was recognised that establishing whether the new assessment techniques would be productive in this manner was beyond the scope of the current research.

Summary of Aims and Objectives

The four aims of the research were stated—briefly, that the research be ethical, an example of good science, be clinically relevant, and represent a significant contribution to knowledge. Associated with these aims, a series of objectives were presented that provide a specific benchmark for the evaluation of the success or failure of the research programme.

Summary of Part I

In Part I the foundations of the research were outlined. After the Preface briefly introduced the research, Chapter 1 presented the background of the research in clinical neuropsychology, and having identified general research questions these were reduced to a group of specific, testable, research questions. The limits of the research were outlined. Following from this, Chapter 2 presented the aims and objectives of the research, providing a benchmark for the evaluation of the success or failure of the research programme.

PART II:

NARROWING THE FOCUS

CHAPTER 3: IDENTIFICATION OF TARGET GROUP: CLINIC FILES REVIEW[2]

Introduction

This study was a retrospective review of a cohort of those referred to the Psychology Clinic, Massey University for neuropsychological assessment following traumatic brain injury. The objective of this review was to evaluate the extent to which the assessment process conducted with people who have experienced traumatic brain injury is impeded by physical and sensory disabilities. This was considered a necessary starting place before embarking on an attempt to define more appropriate assessment procedures.

Recognising that individuals with physical and sensory disabilities present a broad range of specific assessment requirements, the second objective of this review was to identify a specific disability group that would become the focus for the remainder of the research. Criteria defined for the selection of an appropriate group were that:

- individuals with these disabilities represent a reasonable proportion of the traumatically brain injured population, in order that a sample of sufficient size could be recruited for the clinical phase of the research,
- standard administration techniques had been shown to be inadequate with people with these disabilities, and
- there was information to suggest that development of alternative assessment procedures would prove useful for such a group.

For the purposes of the current study, a standard formulation of what a comprehensive assessment is comprised of was derived from Lezak (1995). It was similar to that used in recent large multi-centre investigations (e.g., Adams & Heaton, 1990; Butters et al., 1990; Kreutzer et al., 1993), and encompassed general intelligence, learning and memory, attention, language, visuoperception and visuoconstruction, motor function, information processing speed, and higher executive functioning, as well as personal, emotional and interpersonal functioning.

---

[2] The material presented in this chapter formed the basis of Babbage and Leathem (1997).

The core measures used for the assessment of each of these domains, during the period reviewed, are presented in Table 1. These measures had been selected on the basis that they had acceptable psychometric properties, were endorsed in the research literature, and could be administered in the time typically available for an assessment. In accord with the outcomes after traumatic brain injury reported in the literature, greater focus had been placed on learning and memory. This set of core measures continues to evolve and since the period examined in the review other scales have been adopted including the third editions of the Wechsler scales (Wechsler, 1997a).

It was expected that for people with physical and sensory disabilities some assessment would be possible using standard administration of measures. However, it was hypothesised that these core measures would prove to be inadequate as a comprehensive assessment for the individuals with physical and sensory disabilities identified in this review, and that many alternative, adapted and ad hoc measures would have been used. In addition, it was expected that there would be some domains of functioning which would prove impracticable or impossible to assess in spite of adaptations.

Method

Sample

The files of all individuals with traumatic brain injury assessed at the Psychology Clinic, Massey University, from January 1988 to August 1997 were reviewed. The service is situated in Palmerston North, New Zealand, a city of 70,300 people. Individuals are referred to this service from around the lower North Island of New Zealand, which generally provides post-acute assessment and intervention services for individuals continuing to experience difficulty following traumatic brain injury, after initial rehabilitation efforts have been completed.

Only those who had sustained a brain injury after the age of 15 years, and who prior to this had functioned normally were included in this study. Assessment of individuals with prenatal, perinatal, paediatric, or genetic impairments has its own particular considerations, which were beyond the scope of this study.

Table 1. Core measures used for assessment of traumatic brain injury at the Psychology Clinic, Massey University.

Domain	Reference
General Cognitive Functioning	
Wechsler Intelligence Scale-Revised (WAIS-R)	Wechsler (1981)
Learning and Memory	
Digit Span Forward of WAIS-R	Wechsler (1981)
Rey Auditory Verbal Learning Test	Rey (1964); Taylor (1959)
Logical Memory subtest of Wechsler Memory Scale-R	Wechsler (1987)
Rey-Osterreith Complex Figure recall trials	Rey (1993); Osterreith (1993)
Milner Maze	Bowden & Smith (1994); Milner (1965)
Attention	
Digit Span Backward, Digit Symbol, and Arithmetic subtests of WAIS-R	Wechsler (1981)
Stroop Color-Word Test	Golden (1978)
Paced Auditory Serial Addition Task	Gronwall (1977); Gronwall & Sampson (1974)
Trail Making Test	Army Individual Test Battery (1944)
Verbal Functions and Language	
Vocabulary, Comprehension, Similarities subtests of WAIS-R	Wechsler (1981)
Twenty Questions confrontational naming	Laine & Butters (1982)
Visuoperception and Visuoconstruction	
Picture Completion, Picture Arrangement, Block Design, and Object Assembly subtests of WAIS-R	Wechsler (1981)
Rey-Osterreith Complex Figure copy trial	Rey (1993); Osterreith (1993)
Motor Function	
Finger Tapping Test	Halstead (1947); Reitan & Wolfson (1993)
Information Processing Speed	
Digit Symbol subtest of the WAIS-R	Wechsler (1981)
Stroop Color-Word Test	Golden (1978)
Paced Auditory Serial Addition Task	Gronwall (1977); Gronwall & Sampson (1974)
Higher Executive Functions	
Wisconsin Card Sorting Test	Berg (1948); Grant & Berg (1948)
Twenty Questions	Laine & Butters (1982)
Stroop Color-Word Test	Golden (1978)
Trail Making Test	Army Individual Test Battery (1944)
Personal, Emotional and Interpersonal Functioning	
Symptom Checklist 90-R	Derogatis (1994)
Neuropsychology Behavior & Affect Profile or Patient Competency Rating Scale	Nelson et al. (1989); Prigatano et al. (1986)

In all, 288 files were examined. Over the period examined, 15 individuals were referred twice, and 1 was referred three times, resulting in a total sample of 271 individuals. The sample was composed of 81% Pakeha (New Zealand European), 17% Maori, and 2% other ethnicities. Males made up 70% of the sample. This distribution of gender, and the disproportionately high representation of Maori (c.f. 12% of general New Zealand population), corresponds with previously reported figures after traumatic brain injury (Accident Rehabilitation and Compensation Insurance Corporation, 1993). Age in the sample ranged from 15 to 76 years, with a median age of 30.

A wide range of traumatic pathologies were represented in the sample, including motor vehicle accidents (56%), falls (20%), collisions (12%), and assaults (5%). In this sample, time since injury ranged from 0 years 1 month to 50 years 0 months, with a median of 29 months. Seventy-five percent of the sample were 12 months post-injury or longer. In many cases hospital records, and thus Glasgow Coma Scale scores, were not available, so severity criteria from Lezak (1995, p. 173) based on length of post-traumatic amnesia were employed. On this basis, 39% were classified as having received very severe or extremely severe brain injuries, 20% severe brain injuries, and 23% moderate brain injuries. Eighteen percent had received very mild or mild traumatic brain injury but had re-presented for medical attention due to traumatic brain injury related difficulties.

Procedure

Operational Definition Of ‘Hard-To-Assess’

On the basis of a review of the background information in their files, individuals were also rated for their level of impairment in four areas—motor functioning, vision, speech, and hearing and language comprehension. Past clients were classified for this review as ‘Hard-to-Assess’ (HTA) if they met any one or more of the following criteria:

- Unable to use their dominant hand to write, draw or manipulate, either through hemiparesis, amputation, or quadriplegia.
- Mute or unable to communicate verbally.
- Completely deaf, or unable to comprehend spoken language.

- Blind (including cortical blindness), or with severe perceptual difficulties (such as homonymous hemianopia).

These criteria were selected to include the range of physical and sensory disabilities that would interfere with the process of neuropsychological assessment. Other disabilities, such as lower body paralysis for example, were not included in the criteria as despite being a significant disability they do not interfere with standard neuropsychological assessment.

During the course of the review, a number of individuals were identified as difficult to assess for reasons outside these criteria. Some were experiencing pain to a degree that interfered with test administration, others had psychological problems that conflicted with testing attempts, and some had such reduced attention that testing was abandoned. Finally, a small number of individuals simply refused to complete all measures. While of clinical importance, these difficulties fall outside of the focus of this study and were excluded from the analysis.

Nature of Assessment Conducted

Assessment for each individual was coded as having been conducted either using standard core measures, using alternate measures, using adaptations of core measures, using adaptations of other measures, or using new ad hoc measures. Where assessment was not conducted in a domain, reasons for this were recorded.

Results

Using the stated criteria, 22 individuals (5 females and 17 males) were identified as HTA from the 271 files available, of whom 71% were Pakeha and 29% Maori. They ranged in age from 19 to 73 years. Eleven individuals met the HTA criteria in one of the areas (vision=5, motor=4, speech=2), 7 in two areas (vision-motor=1, motor-speech=5, vision-speech=1), and 4 for three of the areas (motor-vision-speech). No one met the criteria for being HTA because of hearing difficulties. Time since injury ranged from 5 months to 20 years, and severity ranged from moderate (6%) to severe (12%), very severe (41%) and extremely severe (41%). All of these injuries involved motor vehicles—one individual was hit by a car while walking, while the remainder were in vehicles (78%) or on motorbikes (17%) involved in accidents.

Three of these individuals were excluded from further analyses because with these individuals no assessment whatsoever had been possible. One of the individuals in the speech group displayed an autistic-like disengagement from the world around her, and no formal testing proved possible. For two of the individuals in the vision-motor-speech group, psychometric assessment was not possible because reliable communication had not been established. Therefore, 19 individuals remained in the HTA sample.

In analysing the methods used to conduct assessment with these individuals, measures were divided into the eight domains displayed in Table 1 (excluding *General Cognitive Functioning*). Figure 1 displays the proportion of the domains assessed with standard, adapted, and newly created measures for each of these 19 individuals. For each individual, where any newly created or adapted measures were required for the assessment of a domain, that domain was coded as using adapted or new measures.

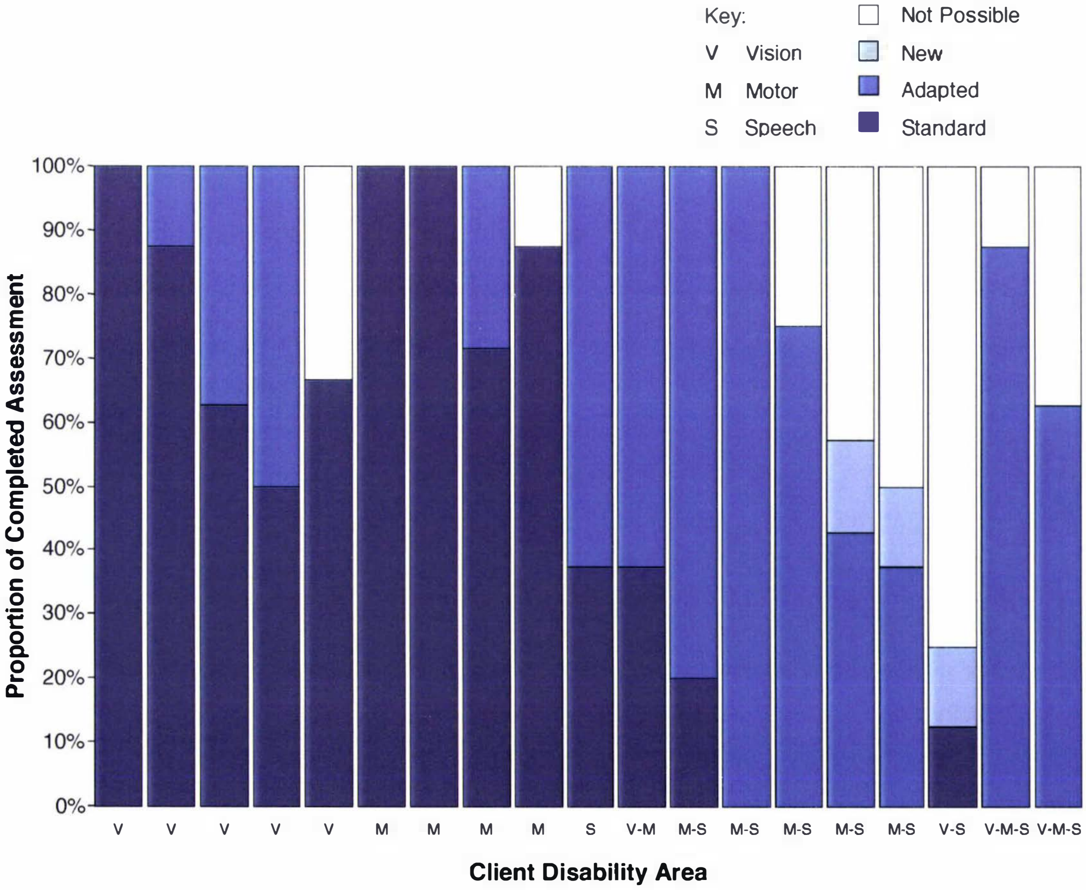


Figure 1. Proportion of assessment completed for each client with standard, adapted and new measures.

In four cases it had not been considered necessary by the assessor to examine particular domains in order to answer the referral questions. This affected between 13% to 75% of the assessment domains of these four individuals, from the Vision, Motor, and Motor-Speech groups. Additionally, in some cases individuals did not complete the full assessment, either because they withdrew their consent for participation, or because they did not return for subsequent sessions. Again, this affected from 13% to 75% of the domains, of five individuals in the Vision, Motor, and Motor-Speech groups. (There was no overlap between these five individuals and the four individuals mentioned above.) In both cases displayed data is for those domains in which assessment was completed, or where it proved to be impossible. Thus, the displayed data are not artificially inflated either by motivational factors of clients or by the omission of measures through the choice of the assessor, but rather reflect genuine areas of assessment difficulty.

An examination of Figure 1 shows that the extent to which adapted and new measures were required, and to which domains could not be assessed, was dependent on disability group. The requirement for adapted or new assessment measures, and the inability to assess some domains, significantly increased with number of disability areas, $\chi^2(4) = 66.14$, $p < .001$. Where individuals had disability in only one area, 85% of the domains were assessed solely with standard administration measures, while 15% were examined using adapted measures. However, for individuals with two areas of severe disability only 19% of domains were assessed with standard measures, while 78% required adapted measures and 3% of domains were assessed partly using newly created measures. For the two individuals who met the criteria for being HTA in three areas, 100% of domains were assessed using adapted measures.

The domains in which adapted and new assessment methods were employed, summarised by group, are listed in Table 2. Where adapted or new measures were used with any member of a group, this is listed in the table. Similarly, Table 2 also shows where assessment of a domain was impossible with some individuals in a group. The table thus provides a ‘worst case’ view of areas in which there may be assessment difficulties with individuals with this range of disabilities.

Discussion

In every group, there were some areas that were not assessed at all. Sometimes this was because the disability made parts of an assessment unnecessary or inappropriate. (For example, there is no need to assess the motor speed of a completely paralysed individual.) Assessment was also curtailed due to assessment constraints (such as in one case time available for a hospital-based assessment), individuals refusing to complete measures (including not returning for a second assessment session), or because a reliable communication method with the individual could not be established. For two individuals in the motor-speech group, assessment was not completed because they were unwilling to participate for the time required for even a standard assessment, highlighting that successful comprehensive assessment is not guaranteed merely by having access to the appropriate test materials.

Table 2. Domains in which assessment difficulties arose as a function of areas of disability.

Domain of Assessment	**Area of Client Disability**						
	Vision (n = 5)	**Motor** (n = 4)	**Speech** (n = 1)	**Vision-Motor** (n = 1)	**Motor-Speech** (n = 5)	**Vision-Speech** (n=1)	**Vision-Motor-Speech** (n = 2)
Learning and memory			**A**	**A**	**N**	**✗**	**✗**
Attention	**A**	**A**	**A**	**A**	**A**	**✗**	**A**
Language			**A**		**✗**	**✗**	**A**
Visuoperception and construction	**✗**			**A**	**A**	**N**	**A**
Motor functioning		**✗**			**✗**		**✗**
Information processing speed	**A**			**A**	**✗**	**✗**	**✗**
Higher executive functions	**✗**	**A**	**A**	**A**	**✗**	**✗**	**A**
General personal, emotional and cognitive functioning	**A**		**A**		**A**	**✗**	**✗**

Note:

Blank	Standard assessment techniques used (from the core battery or otherwise)
A	Areas in which adaptations to measures were required
N	Areas in which new assessment measures were created
✗	Areas in which, with some individuals, no assessment of that domain was possible

This aside, where assessment would have been helpful, 17% of the assessments were not conducted due to the direct effects of the clients' disability. That is, there were no suitable measures, standard, adapted, or ad hoc. Individuals with disabilities in more than one area had the least complete assessments. Contrary to expectations, very few attempts had been made to create ad hoc new measures, with the only attempts made with the motor-speech and vision-speech groups, constituting 3% of their assessment.

Arising from this review was an awareness of the importance of having communication clearly established prior to assessment. In extremely severe cases, communication may be as basic as indicating choice. This can be via vocalisations, eye, arm, or leg movements—any response provided it has been demonstrated to be reliable. The two most disabled individuals in the sample, who had visual, motor, and speech difficulties in combination, proved impossible to formally assess, as a reliable response could not be established. This issue should be seen as quite separate to an examination of cognitive functioning, would often involve other professionals (e.g., speech language therapists) and must precede formal cognitive assessment. A number of authors have written on the establishment of reliable communication with individuals with severe disabilities (for example, see Bigge, 1982, for a detailed and very practical guide).

The basic principle underlying all the adapted measures found in this review was to convert everything to the modes of communication that were available. At its most basic this involved breaking down every measure into a series of choices. Assessments such as this require substantially more time than usual, a precious commodity in health systems where cheaper and more efficient services are often demanded by funding agencies. Therefore, while comprehensive assessment is the ideal, some priority should be given to those assessment areas pertinent to an individual's current needs, even if other domains, for whatever reason, are not assessed.

The 8% of individuals in the sample that were HTA indicates that this is a significant individual group. Additionally, the number of individuals seen is probably an underestimation of those that could be seen—if better assessment were possible. However, no matter how many adaptations are made, some areas may be impossible to assess with direct assessment measures. In these cases, indirect methods could be used instead. Questionnaires for hospital staff and caregivers and interviews with significant individuals can play an important role.

Focus Group for Further Research

The review reported above examined assessment conducted in the past with individuals with disabilities that interfere with the neuropsychological assessment process. The aim in undertaking this review had been to determine the extent to which standard neuropsychological assessment was adequate for the assessment of individuals with physical and sensory disabilities, and to identify an appropriate group on which to focus the remainder of this research project. Criteria for identifying a suitable group were that they represented a reasonable proportion within the traumatically brain injured population, that standard administration techniques had been shown to be inadequate, and that there was information to suggest that alternative assessment procedures might prove suitable for this group.

Based on these criteria the motor-speech group was selected as the focus of the research programme. This group represents nearly a quarter of individuals identified as HTA in the review. Additionally, measures that have been adapted to be suitable for people with motor and speech disabilities would also be appropriate for individuals who have only one of these disabilities. Thus, such adaptations would be suitable for use with 10 out of the 22 HTA individuals, or almost 4% of the traumatically brain injured population reviewed. Of course, measures suitable for people with this level of disability can also be employed with the 92% of people with traumatic brain injuries who do not meet the criteria for being HTA.

Motor and speech disabilities represent a significant impediment to neuropsychological assessment and thus an appropriate focus group for this research. The most common modalities of response for neuropsychological tests are verbal responses, writing or drawing, or manipulating test materials—responses that individuals with Motor-Speech disabilities cannot make. However, this review showed that with adaptations to standard assessment much progress could be made in this group.

Finally, both speech and motor disabilities are *expressive* disabilities, relating to an individual's ability to communicate with and manipulate their environment. In contrast, visual or auditory impairments lead to *receptive* disabilities, impacting an individual's ability to perceive events in and receive information from their environment. Selecting a group with only expressive disabilities greatly simplifies the research question—in what ways can individuals with these disabilities communicate information regarding their responses to neuropsychological assessment stimuli?[3]

In contrast, if certain other groups were selected, such as the Vision-Motor group, both stimulus presentation and response modalities would have to be examined. As experimental control and consistency is always a challenge when working with such complex clinical populations, selection of a target group for the current project that further compromised these factors was considered unwise. Nonetheless, it is hoped that this project may promote future investigation of these questions, providing information relevant to the assessment of individuals with other forms of severe disability.

Summary of Clinic Files Review

Review of the 271 individuals referred following traumatic brain injury revealed 22 (8.1%) who met the criteria for being HTA. Most HTA were those with disabilities in more than one area. Greater number of disabilities was associated with increased need to adapt measures, to develop ad hoc alternatives and to abandon testing altogether. This review of files suggests that assessment needs to be flexible enough to adapt standard measures, for as hypothesised, the core measures, administered in

---

[3] Throughout the remainder of this dissertation, the term *expressive disabilities* will be used to refer to individuals who are unable to communicate verbally (or who can communicate verbally to a limited extent only), and who due to motor disabilities are unable to write, draw, or manipulate test materials. The term, not in common use in the literature, was selected as it clearly highlights the distinct neuropsychological assessment requirements of individuals with this type of disability, as compared to those with *receptive disabilities* (e.g., people who are visually impaired, or hearing impaired).

standard form, were inadequate for the assessment of HTA individuals. Contrary to expectations, very few ad-hoc measures were developed to assess the individuals.

In spite of using alternative measures and adaptations, many domains of cognitive functioning were not adequately assessed—17% of the total assessment reviewed, and up to three-quarters of the assessment of particular individuals. This finding highlights the need for further research to identify possible alternative assessment measures, and where necessary to develop new assessment techniques.

Individuals with expressive disabilities were selected as the focus group for the remainder of the research.

CHAPTER 4: ASSESSMENT DOMAINS IN RESEARCH: JOURNAL REVIEW

Introduction

On the basis of the retrospective review of files from the Psychology Clinic, the group of people with expressive disabilities was selected as the focus for the research. The research questions related to whether a comprehensive assessment across the domains of cognitive functioning could be undertaken with people in this group. To answer these questions, the concept of a *comprehensive* cognitive assessment must be defined.

Clarification of just what constitutes a comprehensive assessment forms the basis of this chapter. Three sources of information were consulted in answering this question. Initially, one of the leading texts of the field, Lezak (1995), was examined for a basic formulation of the domains that constitute cognitive functioning. Next, this formulation was matched against the domains into which researchers divided cognitive functioning in their published research studies. Finally, a survey was conducted to gain the views of clinicians. This survey is presented in the following chapter.

The objective of this study is to provide quantitative information on the domains into which cognitive functioning is divided. This information will be used to ensure that measures selected for the clinical phase of the research cover as broad a range of functioning as possible. Firstly, however, a discussion of the background to the issue is required to put this study in its appropriate context.

Background

There is general consensus that neuropsychological assessment should be broad-based both to provide an overall picture of an individual's functioning, as well as to guard against the risk of a confirmatory bias (Wedding & Faust, 1989). In forensic settings, where the assessment conducted by a neuropsychologist can come under intense scrutiny, assessment should not only be comprehensive, but the procedures and instruments used should be those that are endorsed by the majority of other practitioners in the field (Guilmette & Giuliano, 1991).

Just what constitutes comprehensive assessment, though, varies with several factors, including the population being studied, time available, purpose of the assessment (e.g., clinical, research), and context of the assessment. Further, when the focus moves from general assessment of functioning to focus on specific domains (e.g., in research studies), these are frequently partitioned into increasingly smaller discrete elements. For example, Sbordone (1992) divided attention and concentration into 12 distinct subcategories; Levin et al. (1997) divided executive functions in children into 12 categories; and Francis, Fletcher, Rourke, and York (1992) divided children's motor, psychomotor and visual-spatial functioning into 5 separate factors. However, such fine-grained analysis of specific domains may not be possible or even necessary in all research projects.

In a clinical setting, there appears to be no easy answer as to what constitutes 'adequate' neuropsychological assessment, with much depending on whether the question is focused on the measures used, or on the cognitive domains covered. There is agreement that the measures selected should be relevant to an individual's suspected areas of difficulty, focusing on the prevailing deficits of specific conditions such as AIDS (Butters et al., 1990), or traumatic brain injury (Clifton, Hayes, Levin, Michel, & Choi, 1992). Initial general assessment can be followed up later with investigation of specific areas of difficulty. For example, initial screening for attention and concentration difficulties could be followed with specific measures of elements of attention if required. This approach assumes that while attention and concentration can be separated into a number of elements, there is an underlying meaningful factor that a general measure could examine. This is not a trivial assumption; if there is not a higher-order factor that these elements belong to, then a 'general screening' measure is either a theoretically meaningless combination of unrelated items, or in fact assesses only one aspect of the domain. Selection of assessment measures should always be based on a theory of what these measures examine. (However, the extent to which we know what various tests measure has been questioned; e.g., Dodrill, 1997.)

It has been argued that a 'core battery' of measures should be developed, to be used by all practitioners as a basis for further flexible testing (Benton, 1992; Levin, 1994;

Parsons, 1991). Review of articles proposing such core batteries reveals considerable variation in the measures selected to examine certain cognitive domains (e.g., see a basic battery in Lezak, 1995, p. 122; c.f. Butters et al., 1990; and Kreutzer, Gordon, Rosenthal, & Marwitz, 1993). Variation is partly attributable to different assessment procedures being selected depending on the client group (e.g., assessment of early AIDS-related cognitive changes will be different to that for late stage AIDS-related dementia, Adams & Heaton, 1990; mild vs. severe traumatic brain injury, Acimovic, Keatley, & Lemmon, 1993). In recent years the flexible battery approach to assessment has become increasingly popular, and is now endorsed by 69% of practitioners in the United States, compared with only 15% endorsing a standardised battery approach (from unpublished 1999 survey data, J. Sweet, personal communication, 20 August 1999; see also Sweet & Moberg, 1990; Sweet, Moberg, & Westergaard, 1996). While a core battery of measures may often be used, in theory the domains examined and the measures employed are tailored to the needs of each client and situation (for a discussion see Levin, 1994; Lezak, 1995; Parsons, 1991). Whether this occurs in practice has been questioned (Puente, 1992). However, Guilmette and Giuliano (1991) have suggested that although there are variations in the tests used by practitioners, there is considerable overlap in the domains or functions these different measures actually assess.

If intelligence is what the tests of intelligence measure (Boring, 1923), perhaps in a similar way the domains of cognitive functioning are ‘whatever neuropsychological tests measure’? Along this vein, factor analytic studies provide important information about the construct validity of neuropsychological tests. Various researchers have taken groups of commonly used neuropsychological measures and subjected them to factor analysis, to examine whether the factor structure uncovered matches the variables that test authors and previous researchers have attributed to the tests (e.g., Francis, Fletcher, Rourke, & York, 1992; Heilbronner, Buck, & Adams, 1989; Moehle, Rasmussen, & Fitzhugh-Bell, 1990; Peach, 1992; Robinette, Sherer, & Adams, 1993). However, while such analysis may identify that measures do not have the postulated structure, and indicate other factor structures instead, it can give no guide as to whether these are the domains that should be included in an assessment.

A complimentary approach is to take a top-down perspective, allowing theory to direct assessment and test development. In one sense this is saying that the domains of cognitive functioning are 'whatever neuropsychologists measure'. The distinction between different domains of cognitive functioning in current neuropsychological measures is a result of a slow evolution of the collective understandings of practitioners and researchers, derived partly from clinical experiences, and influenced by other fields including neurology, psychiatry, clinical psychology, and cognitive science. The question of exactly which general domains of cognitive functioning should be examined is being answered by ad hoc consensus of experts based on a broad knowledge of literature, the views of peers, particular features of the client population, and (perhaps) individual quirks in viewpoint.

In summary, a review of the literature suggests that there is limited agreement not only on what measures a core battery should include, but also on the areas that a neuropsychological assessment should cover. In many cases, the specific cognitive domains and processes being examined simply reflect the consensus of the authors as being those that must be assessed for that client group, without an explicit empirical or theoretical basis for the domains selected.

The Current Study

This study quantitatively examines the issue of what researchers view as the domains that constitute human cognitive functioning. The study examines researchers' behaviour in terms of their core output—research journal articles. In this sense, the study examines researcher behaviour through the filter of the editorial process. It was hypothesised that all the domains assessed in the articles could be summarised into a group of common domains, and as the literature describes the consensus of researchers, that these domains would constitute (or at least closely approximate) the ideal 'adequate' neuropsychological assessment. This would provide part of the basis for selecting the domains of functioning for which tests would be developed for individuals with expressive disabilities. It was further hypothesised that, while there would be variation in measures used by researchers, certain tests in each domain would be widely used compared to the alternatives, indicating their general acceptance by researchers as suitable measures of that domain.

The study undertaken was a review of all articles that appeared in four journals dedicated to neuropsychology in 1997. The review was focused on the domains into which researchers divided cognitive functioning. While assessments for particular populations may focus on certain areas (such as memory in traumatic brain injury, or verbal functions after stroke), these presumably would be subsets of the overall domains of cognitive functioning. For this reason, while variation between population groups was expected, analysis was not restricted to one population.

Method

All issues of *Neuropsychology (NP),* the *Journal of Clinical and Experimental Neuropsychology (JCEN)*, the *Journal of the International Neuropsychological Society (JINS)*, and *Archives of Clinical Neuropsychology (ACN)* published in 1997 were reviewed. Excluding abstracts of conference presentations and book reviews, 250 articles were published in these journals during the year. A further 69 of these articles were excluded as they were theoretical treatments of a topic (n=11), research reviews (n=19), dealt with clinical practice issues (n=3) or were factor analyses of specific measures (n=7)[4]. Papers that presented empirical research that did not involve assessing the cognitive functioning of a group of individuals were also excluded (n=8). As research, rather than clinical practice, was the focus of the current investigation, case studies were excluded as well (n=10). Finally, studies were excluded that analysed the results of a variety of neuropsychological tests in a population, but did not define what the authors believed these tests to measure (n=11). One hundred and eighty one articles were thus included for final analysis (*NP*=53, *JCEN*=53, *JINS*=37, *ACN*=38).

Lezak (1995) divided the tests in the compendium of her text into 9 areas—Orientation and Attention, Perception, Memory, Verbal Functions and Language Skills, Construction, Concept Formation and Reasoning, Executive Functions and Motor Performance, Tests of Personal Adjustment, and Tests for Functional Complaints. As the most widely used and comprehensive reference text in the area of neuropsychological assessment, this list was considered, with one

---

[4] However, factor analytic studies of *domains* of cognition were included.

exception, to be an acceptable yardstick against which to reference the domains assessed in the journal articles reviewed. Lezak (1995) presented executive functions and motor performance in one chapter, but essentially treated them separately. As initial examination of the literature suggested that executive functions and motor performance were universally treated independently, these two domains were separated.

Many articles presented findings from certain measures that were a subset of a larger battery, with other portions reported elsewhere. In such cases if a study reported summary statistics about a test as a type of demographic information in the group being studied (e.g., memory score ranges for each group reported along with age ranges and gender breakdown), but did not analyse these data as part of the study, that study was not recorded as having assessed that area of functioning.

Results

Demographic breakdown of the articles reviewed is presented in Table 3. Articles assessed all age groups, and both healthy individuals and a wide variety of clinical populations. Due to variation in the age ranges included in various studies, it was not possible to categorise older age groups, into 'Young Old' and 'Old Old', for example. Instead, a general 'Elderly' category is reported for all articles which examined over 65 year olds. Where populations were examined both below and above this figure (e.g., 25-35 year olds and 65-75 year olds compared), articles were recorded as 'Adult and Elderly'.

While first authors of articles were from 16 countries, 71% of the research was carried out in the USA, and 78% in North America. Articles assessing healthy individuals made up the largest group (n=47), followed by traumatic brain injury (n=19) and Alzheimer's disease (n=15). Sixteen articles examined multiple populations (e.g., Korsakoff's, anoxic and encephalitic groups examined in one article).

Table 4 presents the proportion of articles assessing each of the 10 domains of functioning derived from Lezak (1995). In each case, an article was counted if it assessed any aspect of a domain (e.g., verbal memory in 'Memory and Learning'). All domains were assessed by a number of articles, with twelve articles assessing Construction, and Concept Formation and Reasoning, the two areas of functioning least examined.

Table 3. Breakdown of articles by age group, population, and country of first author.

Age Group	**Articles**	**Population**	**Articles**
Adult	109	Healthy Individuals	47
Elderly (> 65yrs)	43	Traumatic Brain Injury	19
Children	18	Alzheimer's Disease	15
Adult & Elderly	11	Other Dementia	12
		Medical Patients	11
Country	**Articles**	Stroke	10
		Human Immunodeficiency Virus	7
		Parkinson's Disease	7
United States of America	128	Epilepsy	6
Canada	14	Schizophrenia	4
Australia	8	Multiple Sclerosis	3
United Kingdom	7	Huntington's	3
Netherlands	5	Neurotoxic Exposure	3
Italy	4	Attention Deficit Hyperactivity Disorder	2
Sweden	3	Substance Abuse	2
Israel	3	Multiple Populations in One Study	16
Other European	7	Other Populations	14
Asia/Pacific	2		

Table 4. Percentage of articles assessing each domain of cognitive functioning.

Domains from Lezak	**%**	**Other areas of functioning**	**%**
Memory and Learning	63	Overall Cognitive Functioning	32
Attention and Orientation	28	Information Processing Speed	7
Verbal Functions and Language Skills	26	Premorbid Cognitive Functioning	5
Executive Functions	23	Malingering	4
Perception	18	Laterality	3
Motor Performance	16	Reaction Time	2
Personal Adjustment	14	Arithmetic Ability	2
Functional Complaints	9	Academic Achievement	1
Construction	7		
Concept Formation and Reasoning	7		

In addition to these areas, four other areas of functioning were assessed by more than 5 studies (see Table 4). Almost one third of articles reviewed included some measure of global level of cognitive functioning. This definition included articles where Full Scale IQ, Verbal IQ, and Performance IQ indices of the Wechsler Intelligence Scales (WIS; Wechsler Adult Intelligence Scale-Revised, Wechsler, 1981[5]; Wechsler Intelligence Scale for Children-III, Wechsler, 1991) were used as variables in the analyses, and those which included a general measure designed to assess dementia severity, such as the Mini Mental State Exam (MMSE; Folstein, Folstein & McHugh, 1975).

Information Processing Speed was seen as important in its own right by a number of authors, and was listed as a separate domain of cognitive functioning by as many authors as assessed Construction, or Concept Formation and Reasoning. Additionally, 5% of articles examined Premorbid Overall Cognitive Functioning using psychometric instruments. Finally, 4% of authors included measures of Malingering. Also assessed by five or fewer studies were Laterality, Reaction Time, Arithmetic Ability, and Academic Achievement.

The breadth of the assessment conducted by studies, that is, the number of domains assessed by each study, is displayed in Figure 2, which reveals that the majority of studies assessed between zero and two of the ten domains derived from Lezak (1995). The figure compares the number of domains examined in articles where only portions of wider research study were being reported compared to studies which reported all measures administered to their participants. It was expected that studies would have reported assessment in fewer domains if they were reporting only a portion of the measures given to their participants. However, while a trend was found in this direction, the difference between the two groups was non-significant ($t(179)=1.79, p > .05$).

---

[5] The Wechsler Adult Intelligence Scale-III (WAIS-III, Wechsler, 1997) and the Wechsler Memory Scale-III (WMS-III, Wechsler, 1997) were released in the year the articles under review were published. Due to the delay between research being conducted and being published, none of the studies reviewed used the newer version of these measures.

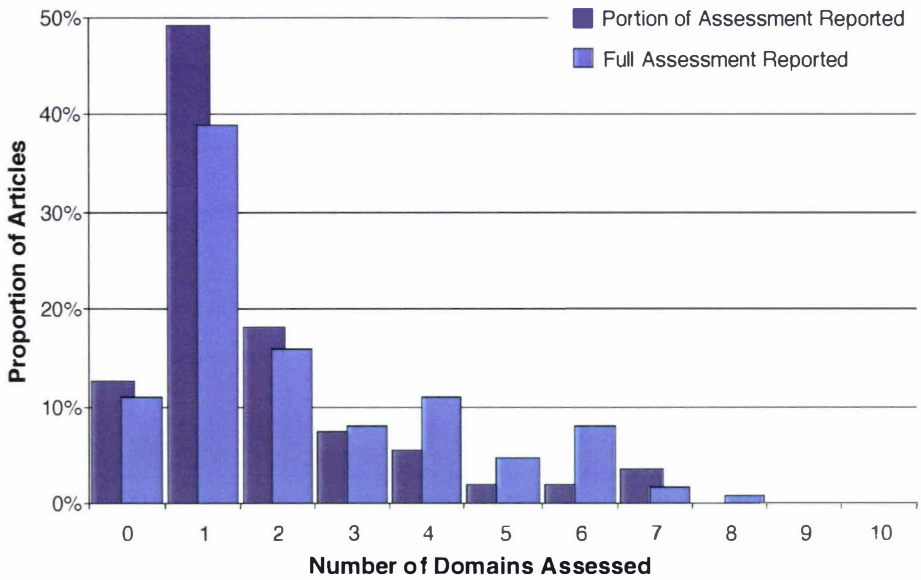


Figure 2. Number of domains assessed in an article as a function of portion of assessment reported.

For each of the 14 domains assessed by more than 5 studies, the measures most commonly used by authors were examined, and are listed in Table 5. For these domains, all measures used by more than one study to assess an area of cognitive functioning were included. In major scales incorporating a number of subtests (e.g., WAIS-R, WMS-R), subtests were treated separately for the purpose of the present analysis.

The most popular measures of Memory and Learning were word list learning tests, with the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987), and the Rey Auditory Verbal Learning Test (Rey, 1964; Taylor, 1959) ranked first and second, and 12 other studies using variations on word list learning tests (2 studies used the CVLT 9-item dementia version and 10 studies used custom word list learning tests). A variety of other well known measures, particularly subtests of the Wechsler Memory Scale-Revised, were used by a group of studies.

Table 5. Number of studies using each measure in each domain.

Memory and Learning	Studies
California Verbal Learning Test	20
Rey Auditory Verbal Learning Test	13
Custom Word List Learning test	10
Rey-Osterreith Complex Figure Test	10
Digit Span (WIS/WMS)	10
Logical Memory (WMS-R)	10
Visual Reproduction (WMS-R)	6
Wechsler Memory Scale-Revised	5
Corsi Block Tapping	4
Verbal Paired Associates (WMS-R)	4
Benton Visual Retention Test	3
Warrington Recognition Memory Test Faces	3
Delayed Recall Index (WMS-R)	3
General Memory Index (WMS-R)	3
Verbal Memory Index (WMS-R)	3
Visual Paired Associates (WMS-R)	3
Autobiographical Memory Interview	2
Benton Visual Recognition	2
CVLT 9-word dementia version	2
Consonant Trigrams	2
Figure and Story Memory Tests	2
Paired Associate Learning	2
Selective Reminding Scale	2
Visual Reproduction (Russell; WMS-R)	2
Visual Memory Index (WMS-R)	2
Other Test/Custom Test	143

Attention and Orientation	
Digit Span (WIS/WMS)	13
Stroop Color-Word Test	8
Trail Making Test	7
Paced Auditory Serial Addition Test	5
Digit Symbol (WIS)	4
Continuous Performance Test	3
Attention/Concentration Index (WMS-R)	3
Digit Vigilance	2
Seashore Rhythm Test (HR)	2
Speech Sounds Perception Test (HR)	2
Symbol Digit Modalities Test	2
Arithmetic (WIS)	2
Other Test/Custom Test	57

Verbal Functions and Language Skills	
Boston Naming Test	17
Custom Category Fluency test	8
Vocabulary (WIS)	8
Custom FAS-type Fluency test	6
Controlled Oral Word Association Test	4
Token Test	4
Complex Ideational Material subscale (BDAE)	2
National Adult Reading Test - UK/USA	2
Peabody Picture Vocabulary Test-Revised	2
Wechsler Intelligence Scales: Information	2
Reading (WRAT-T)	2
Woodcock Reading	2
Other Test/Custom Test	48

Executive Functions	Studies
Controlled Oral Word Association Test	16
Wisconsin Card Sorting Test	13
Trail Making Test	11
Stroop Color-Word Test	10
Custom FAS-Type Fluency Test	7
Custom Category Fluency Test	6
Benton Verbal Fluency Test	4
Halstead-Reitan: Category Test	4
Design Fluency	2
Graphical Sequence Test	2
Tower of London	2
Twenty Questions Test	2
Block Design (WIS)	2
Other Test/Custom Test	26

Perception and Construction	
Block Design (WIS)	12
Judgement of Line Orientation Task	5
Developmental Test of Visual Motor Integration	4
Draw-a-clock	4
Fingertip Number-Writing Perception (HR)	4
Rey-Osterreith Complex Figure Test	4
Picture Completion (WIS)	3
Visual Reproduction (WMS-R)	3
Benton Visual Retention Test	2
Other Test/Custom Test	48

Motor Performance	
Finger Tapping Test (HR)	10
Grooved Pegboard	9
Trail Making Test	7
Grip Strength/Hand Dynamometer (HR)	6
Developmental Test of Visual Motor Integration	2
Purdue Pegboard	2
Digit Symbol (WIS)	2
Other Test/Custom Test	14

Personal Adjustment	
Beck Depression Inventory	9
Minnesota Multiphasic Personality Inventory-II	6
Hamilton Rating Scale	5
Geriatric Depression Scale	4
State-Trait Anxiety Inventory	4
Symptom Checklist-90-Revised	2
Other Test/Custom Test	19

Functional Complaints	
Mini Mental State Exam	2
Vineland Adaptive Behavior Scales	2
Other Test/Custom Test	18

Concept Formation and Reasoning	
Similarities (WIS)	4
Judgement (NCSE)	2
Similarities (NCSE)	2

(table continues)

Concept Formation and Reasoning (contd.)	
Raven's Colored Progressive Matrices	2
Wisconsin Card Sorting Test	2
Other Test/Custom Test	12
Overall Cognitive Functioning	
Full Scale IQ (WIS)	18
Verbal IQ (WIS)	12
Performance IQ (WIS)	9
Mini Mental State Exam	8
Vocabulary (WIS)	7
Block Design (WIS)	6
Dementia Rating Scale	5
Information (WIS)	3
Similarities (WIS)	3
Bayley Scales of Infant Development	2
Mental Development Index (BSID)	2
McCarthy Scales of Children's Ability	2
Shipley Vocabulary	2
Comprehension (WIS)	2
Object Assembly (WIS)	2
Other Test/Custom Test	26
Information Processing Speed	
Stroop Color-Word Test	4
Symbol Digit Modalities Test	4
Information Processing Speed (contd.)	
Trail Making Test	4
Digit Symbol (WIS)	4
Paced Auditory Serial Addition Test	2
Digit Span (WIS/WMS)	2
Other Test/Custom Test	1
Premorbid Cognitive Functioning	
National Adult Reading Test: UK/USA	4
Vocabulary (WIS)	2
Other Test/Custom Test	6

Note:
Abbreviations used in the table:

WIS	Wechsler Intelligence Scales (WAIS-R or WISC-III)
WMS-R	Wechsler Memory Scale-Revised
HR	Halstead Reitan
BDAE	Boston Diagnostic Aphasia Exam
BSID	Bailey Scales of Infant Development
CVLT	California Verbal Learning Test
NCSE	Neurobehavioral Cognitive Status Exam
WRAT-R	Wide Range Achievement Test-Revised

However, in the 114 studies that assessed Memory and Learning, 143 measures were either used in only 1 of the 114 studies, or were custom measures created for a study.

Fifty-one studies assessed Attention and Orientation, with the most popular measures being Digit Span (25%; Wechsler, 1981), Stroop Color-Word Test (16%; Golden, 1978), Trail Making Test (14%; Army Individual Test Battery, 1944) and Paced Auditory Serial Addition Test (10%; Gronwall, 1977). Again, measures that were used in only one study or that were custom-developed for a study featured highly, with 57 such measures being used in the 51 studies.

The Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983) was the most popular measure used to assess Verbal Functions and Language Skills, with 17 of the 47 studies using this measure (36%). Verbal fluency tests of both letter- and category-fluency dominate the list after this, with the WIS Vocabulary (17%) and the Token Test (9%; Boller & Vignolo, 1966; De Renzi & Vignolo, 1962) being the only other measures used in more than 5% of the studies. Across the 47 studies there were 48 measures that were used in only one study, or were custom developed for a study.

Over the 42 studies that assessed Executive Functions, 26 tests were used by only one study or were custom-developed. The Controlled Oral Word Association Test (Benton & Hamsher, 1989) was used by the greatest proportion of studies (38%), followed closely by the Wisconsin Card Sorting Test (31%; Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The two other most commonly used measures of Executive Functions were the Trail Making Test (26%) and the Stroop Color-Word Test (24%)[6].

Perception and Construction were initially considered separately in this study. However, an examination of the measures used by the 33 studies that assessed some aspect of Perception, and the 13 studies that assessed Construction, revealed that the majority of tests included by some authors under Construction (or 'Visuoconstruction' or 'Visuospatial') were the same as those included by other authors under Perception (or 'Visuoperception' or other labels). Therefore, for the purposes of this analysis, Perception and Construction were combined. Thirty-nine articles assessed Perception, Construction, or both. While the majority of articles assessing Perceptual abilities were examining visuoperception (8 of the 9 tests used by more than one article involve visuoperception), other aspects of perception were also examined. The only perceptual test not including visuoperception that was used by more than one article was the Halstead-Reitan Fingertip-Number Writing task, used by four researchers (10%). The most frequently employed test in this domain was the WIS Block Design subtest, used in 31% of the articles. Again, however, many articles employed tests used by no other article, or custom-developed for their study (48 measures in 39 articles).

---

[6] In her compendium, Lezak (1995) included measures of verbal fluency under *Verbal Functions and Language Skills*, although she also noted the relationship between verbal fluency and frontal lobe functioning. Studies examined here variously included measures of verbal fluency under a *Verbal Functions and Language Skills* factor, an *Executive/Frontal Lobe Functioning* factor, or in a separate *Verbal Fluency* factor. Due to the relationship of these measures to frontal lobe functioning, where a study included a separate *Verbal Fluency* factor it was included here under *Executive Functions*.

That measures of verbal fluency were commonly included also in the *Verbal Functions and Language Skills* domain highlights the extent to which actual measures may straddle the conceptual boundaries between theoretical domains of cognitive functioning. Very few tests, with the possible exception of simple motor tasks, can be considered to assess a single domain of functioning independently from other abilities.

The most frequently used tests of Motor Performance were the Halstead-Reitan Finger Tapping Test (36%; Halstead, 1947; Reitan & Wolfson, 1993), the Grooved Pegboard (32%; Matthews & Kløve, 1964) and the Trail Making Test (25%). Fourteen measures of motor performance were used by only one article, or were custom designed for one of the articles in the study.

Measures of personal adjustment were used in 25 of the articles. Three of the top four were measures of depression (Beck Depression Inventory, Beck, 1987; Hamilton Rating Scale, Hamilton, 1960, 1967; and the Geriatric Depression Scale, Yeasavage et al., 1983) with the MMPI-II (24%; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the State-Trait Anxiety Inventory (16%; Speilberger, 1983) also being frequently used.

Very little consistency was found in the measures used by researchers to assess Functional Complaints, with the 17 articles that assessed this domain using 20 different measures. Similarly many different measures were used to assess Concept Formation and Reasoning, with 17 different measures being used over the 12 studies that examined this area.

The Wechsler Intelligence Scales Full Scale, Verbal, and Performance IQ scores were the most common measures used of Overall Cognitive Functioning in the 58 studies that examined this area, being used in 31%, 21% and 16% of the articles, respectively. Individual subtests of the WIS were also widely used as indices of general cognitive functioning. Twenty-six different measures were used by only one study, or were designed specifically for a study.

Information Processing Speed was assessed by 13 articles, with WIS Digit Symbol, the Symbol Digit Modalities Test (Smith, 1982), the Stroop Color-Word Test, and the Trail Making Test each being used by 4 of the articles (31%). Eighteen other tests were used, including 16 different tests used by only one article, or developed for a particular study.

Only nine articles examined Premorbid Cognitive Functioning. Four utilised the National Adult Reading Test or its American version (44%; Nelson, 1982; Blair & Spreen, 1989), two used the WIS Vocabulary subtest (22%), and 6 other measures were each used in only one article. Finally, Malingering was examined by eight articles. These articles used 11 different measures of Malingering, and all measures were used by only one study.

Discussion

The hypothesis that researchers would tend to group cognitive functioning in similar ways appears to be supported by this study. No researcher(s) assessed all the areas of functioning defined. While the theoretical distinction between perception and construction is sound and should not be blurred, for the purposes of this study they were combined as they often appeared to be treated interchangeably as domain labels, with aspects of the same measures being used to assess these two domains.

In addition to the domains derived from Lezak (1995), Information Processing Speed emerged as a domain assessed separately by 7% of the studies. This may be an area warranting investigation in a greater number of populations and studies. It is interesting to note that, along with Working Memory, Processing Speed is one of the two new factors which the third edition of the Wechsler Adult Intelligence Scale added to the domains that it assesses (Wechsler, 1997a, p.13). Laterality, Reaction Time, and Arithmetic Ability were assessed by a small number of studies. These areas of investigation do not seem to be encompassed by the other domains defined. In a comprehensive assessment of cognitive functioning, these areas would need to be assessed.

Despite suggestions that the time for global indices of cognitive functioning has past (e.g., Lezak, 1988), many articles reported IQ scores for participants as part of their demographic data, and almost one third of articles reviewed in this study used some global measure of cognitive functioning as a variable in their analyses. The temptation to include such an index in analyses is understandable. However, the use of a single subtest of a measure as an approximation of 'Overall Cognitive Functioning' (such as WIS Vocabulary or Block Design) seems questionable in the light of current understanding of the broad range of abilities involved in intelligent behaviour.

It was hypothesised that in each domain there would be one or two tests endorsed by a significant number of researchers. While this hypothesis was partially supported, there was also considerable variation in other tests used. In all domains the number of tests which were used by only one study (some of which were developed for those studies) outnumbered the top ranking measure. This diversity is perhaps both a strength and a weakness. While researchers use many measures to assess the same domain, there is the potential for new insights, from the varied methods of examination. However, with little consistency between measures employed, comparing results between studies and populations is more difficult.

It would appear that, with the addition of Information Processing Speed, the 10 domains of Memory and Learning, Attention and Orientation, Verbal Functions and Language Skills, Perception, Construction, Executive Functions, Concept Formation and Reasoning, Motor Performance, Personal Adjustment, and Functional Complaints capture the domains commonly examined in the assessments of researchers included in this study. In addition to these areas, many authors included a measure of Global Cognitive Functioning, and a number examined Premorbid Cognitive Functioning and Malingering. Arithmetic Ability and Reaction Time were assessed by some authors, and should be considered for inclusion in a research battery when wishing to conduct a broad assessment.

Cognition is a complex construct involving various domains each requiring specific examination, and no one measure can assess all domains. At the same time, while theoretical domains of cognitive functioning may be discrete, most neuropsychological assessment measures are not and straddle two or more areas of functioning. Therefore, in order to be able to interpret results adequately, ideal research would assess a similarly broad range of functions to those examined in clinical practice. In contrast, the current review found that many researchers only included measures primarily associated with one or two domains of cognitive functioning. While the suggestion of broad-based research assessment would have considerable resource implications, increasing the amount of time required from both researchers and participants, the consideration of these issues by researchers designing projects will lead to more robust research findings.

Summary of Journal Review

The current study suggests that, with the addition of Information Processing Speed, the 10 general domains identified from Lezak (1995) appear to capture the breadth of most neuropsychological assessments conducted for research purposes. The next chapter examines issues relating to the adaptation of assessment measures for use with individuals with expressive disabilities, and investigates whether the research formulation from the current study, of the domains of cognitive functioning, is supported by clinicians. A consensus between researchers and clinical practitioners on the domains that constitute an adequate, broad-based, neuropsychological assessment would be a firm foundation on which to depart on an attempt to modify assessment for individuals with expressive disabilities.

CHAPTER 5: EXPERT CONSULTATION: PRACTICE SURVEY[7]

Introduction

It was considered desirable to investigate current best practice when working with individuals with expressive disabilities before embarking on the examination of new assessment techniques. In a well-established area of research, this could often be achieved through a review of the professional literature. The review (see Chapter 1) revealed that only a limited amount has been written to date on the psychometric assessment of individuals with severe physical and sensory disabilities. Therefore, it was necessary to canvas the opinions of a group of expert practitioners regarding their approach when working with people with expressive disabilities. The objectives of the study were to identify any psychometric instruments that this group of expert clinicians considered appropriate for clients with expressive disabilities, and investigate any techniques they found useful for the adaptation of existing measures.

Additionally, as any development of new assessment techniques must be based on a formulation of the domains that should be measured, practitioners were asked about the domains that they considered should constitute a neuropsychological examination. These data were collected to stand alongside the data reported in the previous study, which examined the domains within which researchers formulate comprehensive neuropsychological assessment (Chapter 4).

As this study was exploratory in nature, no specific hypotheses were made. Rather, the aim was to highlight techniques in adapting measures, and issues to be considered in the development of the research protocol for the clinical phase of this research.

Method

Participants

The 17 female and 17 male psychologists who participated in this study ranged in age from 29 to 71, with a median of 46, and a mean of 45.1 (standard deviation

---

[7] Earlier versions of material presented in this chapter formed the basis of Babbage and Leathem (1998b), and Babbage and Leathem (1999a). Information from this study was also discussed in Babbage and Leathem (1998a). See Appendix A.

12.6). Three participants did not report their age. Participants were registered (or in the case of doctoral candidates, seeking registration) in the United States (n=21), New Zealand (n=9), Canada (n=2), and Australia (n=2). Psychologists from the USA were spread across the country, working on the West Coast (n=3), in the South West (n=3), Mid West (n=5), South (n=3) and on the East Coast (n=3). Three respondents did not list the state in which they were registered, and 1 participant was registered in 4 of these 5 areas. Twenty-eight participants had completed doctoral degrees: Ph.D (n=20), Psy.D (n=3), Ed.D (n=2), or doctoral degree plus ABPP Diplomate (n=3). Four of the participants from New Zealand had completed Masters degrees and a subsequent postgraduate diploma (the standard qualifications for clinical training and registration as a psychologist in New Zealand). Finally, 2 participants were current candidates for doctoral qualifications (one Ph.D, one Psy.D).

Time spent working in neuropsychology was examined, as previous research has found (among North American respondents) that the proportion of a clinician's current practice, more than the years they have been practising in the field, is an indication of how well they have been trained (Guilmette, Faust, Hart, & Arkes, 1990). Fifty-nine percent of participants were in full-time active employment in neuropsychology and 32% in part-time employment. Nine percent of the sample were currently in training. On average respondents spent 33% of their time working with individuals with traumatic brain injuries, the focus of the case vignette (presented below).

Materials

The questionnaire used in this study (see Appendix B) examined the approach respondents would take to the assessment of an individual described in a presented case vignette. The vignette, reproduced below, described the fictitious case of an individual who sustained a traumatic brain injury in a motor vehicle accident. In line with the focus group selected for the research (see Chapter 3), the client had disabilities that meant she could not speak, and could only communicate by pointing, or through the use of an alphabet board.

> *"L.M. is a 31 year old female who suffered a closed head injury in a motor vehicle accident where the car she was driving collided with a lamp post.*

> *L.M. was admitted to hospital unconscious, and remained so for 16 hours. Post-traumatic amnesia was estimated at 36 hours, although the effects of medications she was receiving made it difficult to establish this precisely. Since her injury L.M. has been unable to speak, although she appears to understand spoken English. Spinal injuries left her paralyzed from the waist down, with only gross motor control currently regained in her arms. Reliable communication has been established with pointing and the use of an alphabet board with assistance. You have been referred this patient to assess the degree of cognitive impairment, in order that decisions can be made regarding rehabilitation and L.M.'s future directions."*

Respondents were asked to comment on the difficulties of assessing cognitive functioning with this individual, to identify the measures they would employ, and the strategies they would use. They were asked whether they would accept this referral, and if so what modifications they would make when standard administration of their usual tests was not possible.

Participants were asked to select, from a list of assessment measures, the tests that they would use as primary assessment measures in the neuropsychological assessment of an adult patient with closed head injury. They were also asked to identify the secondary measures they would use if their primary measures proved to be unsuitable. These data were collected as a baseline of participants' standard assessment, against which their planned assessment of an individual with physical disabilities could be compared. This list of tests was drawn from those used by neuropsychologists surveyed by Sellers and Nadler (1992). Measures designed exclusively for children or older adults were removed, while a number of recently published measures were added, resulting in a list of 124 tests. Respondents were also encouraged to list any additional measures they might use in such an assessment. In all, respondents endorsed 156 tests and batteries.

As participants were not randomly selected (see *Procedure* below), their opinions could not be assumed to represent the modal neuropsychologist. Therefore, questions were also included which examined participants' positions on such issues as the use of technicians, fixed versus flexible approaches to test selection, and also to determine the proportion of their clinical time spent working with individuals who have suffered traumatic brain injury. These questions were drawn from Putnam and DeLuca (1990), Sweet and Moberg (1990), and Sweet, Moberg, and Westergaard (1996). Participants' methods for assessing an individual with expressive disabilities might vary depending on their approach to assessment in general, so these questions were included to provide a 'snap-shot' of the practice of participants in the sample.

Finally, participants were asked to list the domains of cognitive functioning that, in their opinion, are important to assess in a comprehensive neuropsychological assessment. Secondly, participants were asked the extent to which they viewed it as important to always examine all of these areas in a neuropsychological assessment.

Procedure

Due to the exploratory nature of the study, a random sampling of practitioners was not attempted. The intention, rather, was to elicit the insights of experienced neuropsychologists about assessment of individuals with expressive disabilities. Questionnaires were distributed to two groups of participants. In both groups, participants received return envelopes and questionnaires, which were mailed back to the researchers. Forty-two questionnaires were distributed on an ad hoc basis to neuropsychologists attending the International Neuropsychological Society conference, Honolulu, in January 1998 and a further 23 were directly mailed to neuropsychologists who were personally known to the first supervisor of this research project, and returned anonymously. In all, 34 substantially completed questionnaires (52%) were returned and included in the analysis. This is above the 40% to 50% return rate some authors have suggested as an acceptable level for behavioural research (Kerlinger, 1973).

Results

'Snap-shot': Clinical Practice

Figure 3 displays the proportion of respondents who personally interview patients, who use technicians/psychometricians to collect test data, and who personally observe patients during any or all of the administration of tests. Practices of participants were compared to those of 213 ABPP Diplomates and 207 non-ABPP members of Division 40 of the American Psychological Society (Clinical Neuropsychology; Div40). These data were obtained from Prof. Jerry Sweet (personal communications, 20 August, 9 December 1999) and are drawn from their 1999 practice survey, following up Sweet and Moberg (1990), and Sweet, Moberg, and Westergaard (1996).

As can be seen in Figure 3, all respondents personally interview patients during evaluations, very similar to the 97% and 99.5% reported by Sweet and colleagues for ABPP and Div40 members respectively[8]. Significant differences were found in the use of technicians/psychometricians to collect data, with only 33% of the current sample using technicians, versus 42% for the Div40 sample and 69% of ABPP respondents from Sweet et al., $\chi^2(2)=38.90$, $p < .001$. Finally, 91% of the current sample personally observed patients during any or all of the administration of neuropsychological tests, compared to 86% of Div40 respondents and 77% of ABPP respondents in the Sweet et al. study. These differences were found to be significant, $\chi^2(2)=8.08$, $p < .05$.

The philosophical approach to selecting measures was examined, again compared to the 1999 survey by Sweet and colleagues. Figure 4 shows the proportion of participants endorsing flexible, flexible battery, and standardised battery approaches. The current sample contained a larger proportion of practitioners who endorse a flexible approach to assessment (33% vs. 17% for Div40 and 15% for ABPP) while only 52% endorsing a flexible battery (c.f. 65% for Div40 and 74% for ABPP). Fifteen percent endorsed a standardised battery, similar to the 19% reported for Div40 and 11% for ABPP. The difference in philosophical approaches of the current sample to these other groups was statistically significant, $\chi^2(4)=12.26$, $p < .05$.

---

[8] As almost all participants personally interviewed patients, expected values in other cells were too low to conduct χ^2 analysis.

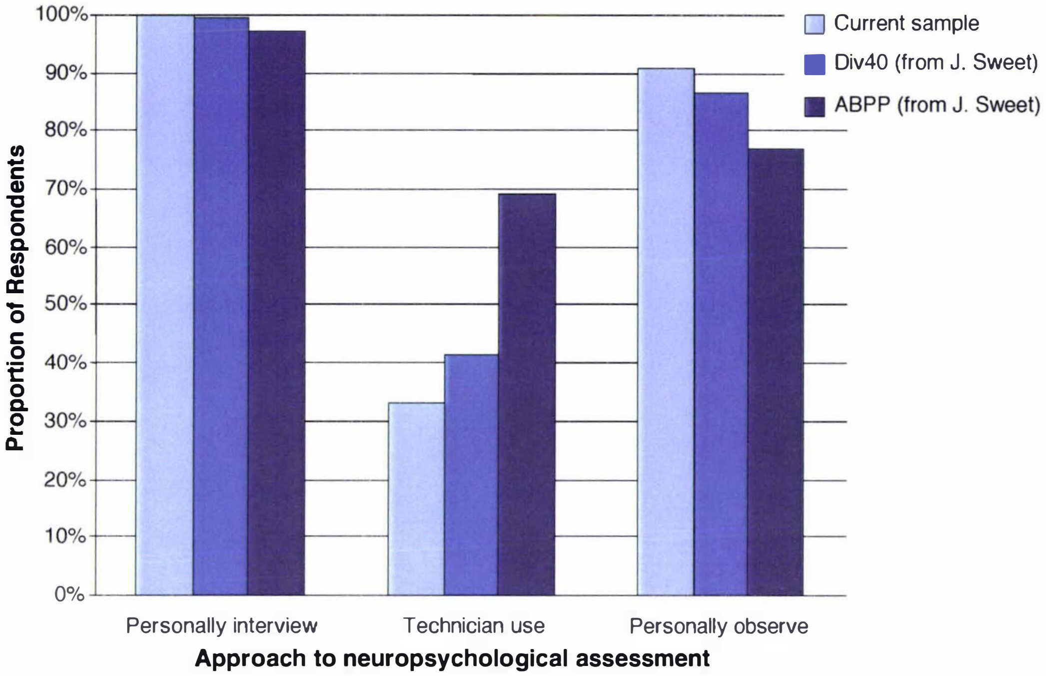


Figure 3. Percentage of clinicians endorsing approaches to assessment in three samples.

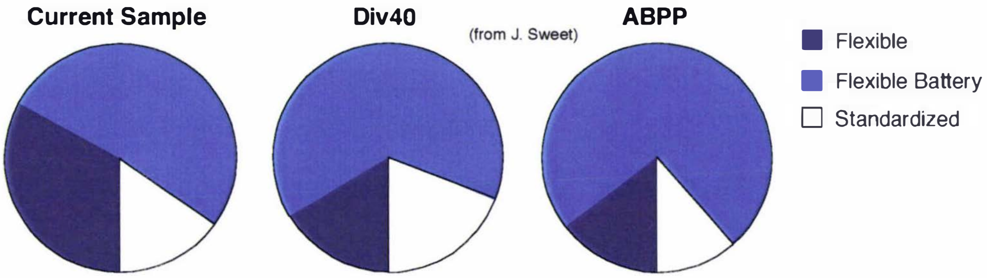


Figure 4. Primary philosophical approach to measure selection.

Case Vignette

Seventy-nine percent of the sample indicated that they would accept the referral described in the vignette. A further 6% of respondents indicated that they would possibly accept the referral providing they felt that the assessment would in fact benefit the person. The majority of respondents indicated therefore that they felt they had the skills and experience necessary to conduct the assessment.

The psychometric measures that respondents would employ in this case were examined. Participants were also asked to indicate those tests from their usual primary and secondary assessment measures that they considered unsuitable for use in this case. As a number of measures were endorsed as appropriate by some participants and inappropriate by others, a simple tally of endorsements was not suitable. Therefore, the ratio of recommendations for versus against each measure was calculated. Measures were only ranked as potentially suitable where they were endorsed by more than 10% of the sample and, where a measure had received mixed ratings, there was at least a 4:1 ratio in favour of the test being suitable. This ratio was selected as a balance between taking account of negative opinions, while not allowing one or two negative responses in the small sample to veto the inclusion of a measure. The 10 measures that met these criteria are listed in Table 6.

Table 6. Rank order of measures recommended for use with L.M.

Measure
Peabody Picture Vocabulary Test
Raven's Progressive Matrices
Vineland Adaptive Behavior Scales
Beck Depression Inventory
Halstead-Reitan Seashore Rhythm Test
Halstead-Reitan Sensory-Perceptual Examination
Benton Line Orientation
Facial Recognition Test
Halstead-Reitan Grip Strength
Hooper Visual Organization Test

Participants considered only 36% of their primary assessment measures to be suitable in this case. While 47% of secondary assessment measures were identified as suitable, with over half of these measures still unsuitable.

It will be recalled that participants were asked to identify the difficulties that might arise in assessing this patient, and the particular strategies that they would use in this case. The information elicited in this way was unstructured and therefore interpreted qualitatively.

All participants recognised that the assessment of a patient with this level of disability could not be conducted solely though standard techniques and procedures. Two-thirds of respondents identified motor control difficulties, and compromised expressive language, as presenting particular blocks to using standard techniques. A further one third indicated that due to motor slowing and the additional demands when tests are modified for use with an alphabet board, results could not be directly compared to norms.

Careful assessment of the communication abilities of the client was recommended by two respondents as an initial step. This would need to include an examination of the reliability of the client's responses, and the extent to which any individuals assisting with communication were able to do so in an unbiased fashion. The possibility of examining other response mechanisms in addition to the alphabet board was also suggested by a quarter of respondents, particularly for simple yes/no questions.

Where an individual performs within normal limits on a test, despite modifications that make the test potentially more difficult (e.g., the longer time required to hold information in memory if spelling it out on an alphabet board), then comparison with norms may be appropriate. However, two participants identified the potential difficulty with dissociating performance due to central versus 'peripheral' deficits, where performance is below average.

Individual respondents identified other difficulties that might be encountered with all individuals—such as the effects of medications, attention and fatigue during the assessment process, and frustration with the assessment—as likely to be particularly exacerbated in this case.

Although respondents had varying opinions on the modification of test materials, and interpretation of these results, the majority suggested that adaptations and modifications of measures would be required in this case. The most commonly suggested modification was to adapt existing tests to a multiple-choice or yes/no format. Answers for these modified measures could be presented on response cards and pointed to by the client, or could be made via the alphabet board. Where timed measures were used these could be altered to allow as much time as required to complete. Two participants discussed the implications of such adaptations on the reliability and validity of the measures—that psychometric properties cannot automatically be assumed to be the same.

In addition to the modification of materials where possible, emphasis was put on the use of collateral information from family members, caregivers, and other professionals working with the client. Other professionals, such as speech language therapists, would be brought into the case if they were not already involved. Respondents indicated that they would also make greater use of behavioural rating scales and observation of the client in her environment, including with her family and in therapy.

Emphasis was placed by some respondents on the need to ensure the referral was in the best interests of the client, and that the assessment was focussed on obtaining information that would be of practical use in her rehabilitation. Other factors also identified in handling this case reiterated other aspects of good practice—consulting other clinicians with experience working with similar individuals, ensuring that test modifications and resulting limitations were clearly identified in the report, and that why such modifications were necessary had been spelled out.

Only one participant felt that assessment could proceed with “minimal difficulty”, if the individual had mastered the use of the alphabet board. In contrast, the majority saw this as a very difficult case and one that would require a substantial amount of individual planning, modification of testing materials and strategies, and potentially a great deal more time.

Domains of neuropsychological assessment

Fourteen domains of functioning were endorsed by more than one respondent as areas important to assess in a comprehensive neuropsychological examination (see Table 7). These include eight specific domains of cognitive functioning, plus assessment of global cognitive functioning, premorbid functioning (cognitive and academic), motor functioning, personal adjustment, and functional complaints. With the exceptions of Concept Formation and Reasoning (31%) and Information Processing Speed (17%), all specific domains of cognitive functioning were endorsed by two-thirds or more of the participants. The only area to receive unanimous endorsement was Memory and Learning. A number of other areas, not listed here, were endorsed by only one participant. All of these areas, although important to the neuropsychological testing process, were outside of the core focus of neuropsychologists (e.g., visual acuity).

Table 7. Percentage of participants endorsing each domain of assessment.

Domain	**%**
Memory and Learning	100
Executive Functioning	90
Perception	79
Attention and Orientation	69
Construction	69
Verbal Functions and Language Skills	66
Motor Performance	66
Global Cognitive Functioning	48
Personal Adjustment	41
Concept Formation and Reasoning	31
Information Processing Speed	17
Academic Achievement	14
Functional Complaints	10
Premorbid Cognitive Functioning	10

Figure 5 displays the number of domains that respondents identified as important to include in a comprehensive assessment. The modal response was 9 domains; however one neuropsychologist named as few as 3 domains, while one identified 12.

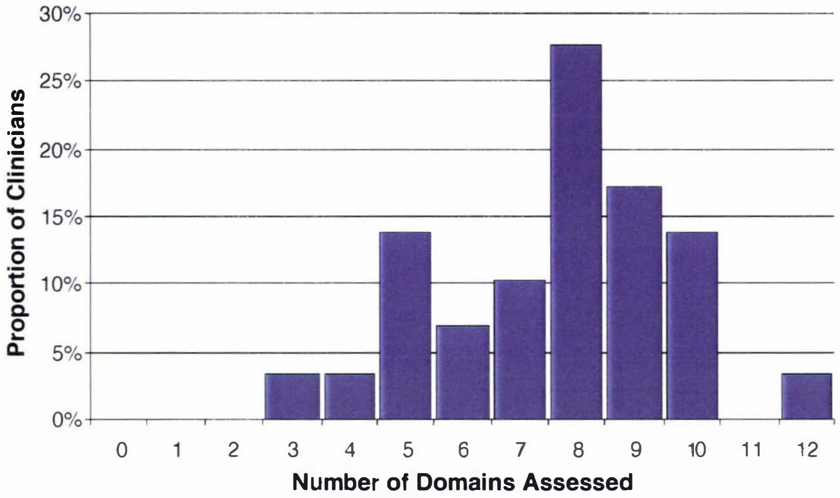


Figure 5. Number of domains named by respondents.

As to whether participants believed the domains of assessment they identified should always be included in a neuropsychological assessment, 13% said, "Yes, regardless of context, patient or referral questions". Permitted to select more than one response, 72% endorsed each of "Depends on referral questions" and "Depends on nature of patient presentation/difficulties", and 38% endorsed "Depends on available resources in assessment context". Only 16% said, "Assessment focus can be directed solely by clinician".

Discussion

Rather than as a survey per se, this study was designed as a 'case consultation' with 34 experienced practitioners, canvassing their views of the nature of a thorough assessment, and examining how they would manage a very difficult assessment. From this investigation, a number of important points were identified.

Domains of Neuropsychological Assessment

Six out of eight of the domains of cognitive functioning were endorsed by over two thirds of participants. It is possible that information processing speed, and concept formation and reasoning, which were identified by fewer participants, are still assessed by many clinicians, and subsumed under other labels. However, there remains considerable variation in the number of domains identified which suggests that there is not necessarily a fixed group of named domains that neuropsychologists would identify as constituting the areas required for an adequate assessment. This fits with the finding that the majority of respondents believe the domains investigated in an assessment would be dependent on a number of contextual factors. Nonetheless, overall the domains identified fit well with the formulation derived from Lezak (1995) and supported in the Journal Review (Chapter 4). The consensus regarding the domains of cognitive functioning from these three sources provided the basis for the domains to be assessed in the clinical phase of the research.

Some researchers have argued that for clinical neuropsychology to inform management and treatment programs it must be continually refining testing procedures to reflect the understandings of clinical neuroscience (e.g., Heinrichs, 1990). The current findings would suggest that, at least in the general domains of cognitive functioning, clinical practice is keeping pace with research.

Assessment of Individuals with Expressive Disabilities

A number of issues were clarified by the survey regarding assessment of people with expressive disabilities. Firstly, regardless of theoretical orientation in selecting neuropsychological assessment measures, selection will have to be substantially modified when assessing an individual with the level of physical disabilities seen in the case described. Among respondents, only 36% of their primary measures and 47% of secondary measures were seen as suitable for use with the individual described.

Secondly, the survey classified the measures most highly ranked by our respondents as being suitable for this individual. These measures provide assessment possibilities in a number of domains of functioning (Attention & Concentration: Seashore Rhythm Test; Language: Peabody Picture Vocabulary Test; Perception: Sensory-Perceptual Examination, Hooper Visual Organization Test, Facial Recognition Test, Benton Line Orientation; Motor: Grip Strength; 'General Cognitive Functioning': Raven's Progressive Matrices; Personal Functioning: Beck Depression Inventory; Functional Abilities: Vineland Adaptive Behavior Scales). In some cases, as with

L.M., further assessment in some areas (e.g., motor, construction) would be ideal but might not be possible. This, of course, would have to be determined based on more information than made available to participants in this study. In addition to these areas, however, memory and learning, concept formation and reasoning, and executive functions stand out as areas requiring further assessment.

Participants made a number of useful suggestions regarding the case vignette, but as significant in participants' responses was the information *not* provided. No respondents identified test batteries or specific measures that they felt would be ideal for use in this case, that would provide a full assessment without substantial modification of assessment techniques. Regardless of their theoretical orientation to test selection, the majority of participants indicated they would accept this referral, and that their assessment would be based on adaptation of existing measures; in some cases, using custom-designed tests was advocated.

There are a number of limitations of the current study. Firstly, the survey methodology is a number of degrees removed from a genuine clinical assessment. Consequently, the amount of information presented in the case vignette is less than might ordinarily be available to neuropsychologists conducting an assessment. Additionally, the data collected here resemble the assessment plan that a clinician might have for an individual with this level of disability, a plan which might be substantially changed by some assessors during the actual assessment. The small sample size and non-random selection of participants in this sample also limits the generalisability of the findings. However, as an exploratory piece of research the study raised some issues helpful in informing the next phase of the research.

The study findings indicate, within the limits discussed above, that for an individual with particular disabilities there will be some domains in which standard assessment is still possible, with few or no modifications to assessment required. In other areas, measures may need to be adapted and suitable caution used in the interpretation of the results relative to normative information. Finally, in some cases it may be that no standard assessment procedures will be suitable and practitioners will have to be creative in their assessment strategies. This study thus provides additional support for the findings from the review of clients seen at the Psychology Clinic, indicating that a diverse group of practitioners would employ similar strategies to those found in that retrospective review.

These studies would suggest that clinicians view 'creative' assessment, interpreted with appropriate caution, as preferable to simply omitting the assessment of a particularly challenging domain of cognitive functioning. While this creativity is a necessary skill and important part of these difficult assessments, the more that standardised measures with appropriate normative information can be employed the better.

Directions for Clinical Phase Protocol Development

The outcome of this study further highlights the need for improvement of the assessment of individuals with expressive disabilities, with much potential for further work developing techniques for assessment, including standardising adaptations to measures. Newer versions of existing measures, such as the Wechsler Scales, are including multiple-choice recognition formats, which allow modifications to be made to standard assessment procedures with some reference to well-designed large normative studies. For example, the WMS-III includes recognition trials, yes/no questions, or both, to supplement the standard administration of subtests, which could provide information about an individual's recognition memory. Unfortunately, normative studies do not report results for individuals who are given *only* the multiple-choice portion of the test, the method of administration most likely to be suitable for those with severe expressive disabilities.

Other scales that do not require complex responses may be amenable to adaptation to an alphabet board, or other form of communication. However, even if an individual is proficient at the use of an alphabet board, this additional requirement is likely to add demands to test requirements. Therefore, rather than using a full length wordlist-learning test such as the Rey Auditory Verbal Learning Test (Rey, 1964; Taylor, 1959) or the California Verbal Learning Test (CVLT, Delis et al., 1987), a shorter version could be used, such as Word Lists from the WMS-III, or the 9-word form of the CVLT (Libon et al., 1996), originally developed for use with older individuals who are already cognitively impaired. Such a shortened form is suggested since, with the added concentration required using an alphabet board, and increased response latency, it is likely that longer versions will prove unnecessarily difficult.

Previous researchers have noted the strong association between visual-perceptual-motor ability and performance on visual memory measures (e.g., Heilbronner, 1992). The majority of non-verbal memory measures require motor responses that would be impossible for most individuals with the level of disability described. In many cases so-called non-verbal memory measures could be converted to have figural recall demonstrated by the client spelling out what they saw via their alphabet board—this possibility indicates, however, that the measures are not purely non-verbal. Normative information for such adaptations is not at this time available, making interpretation difficult. Other measures are available that use a recognition format but extend it beyond a single selection of an earlier presented item to involve more complex memory requirements. For example, the Self-Ordered Pointing Test (Petrides & Milner, 1982; see Spreen & Strauss, 1998) requires individuals to point on each page to a stimulus that they have not previously selected, where pages have the same stimuli arranged in varying positions. This measure thus involves aspects of working memory, organisation of information and self-monitoring. However, this measure is not available commercially and there is only limited normative information. Nonetheless, the concept holds considerable promise for this population.

Due to the essentially multiple-choice nature of the available responses, both the Category Test (Halstead, 1947; see Spreen & Strauss, 1998, for a description of newer versions) and the Wisconsin Card Sorting Test (Heaton et al., 1993) are other measures of aspects of executive functioning that could be used with individuals with this level of disability. It might be necessary for an assistant, or the examiner, to help in manipulation at the direction of the examinee, but the response required to direct this could be made in many ways without necessarily invalidating comparison with standard normative information in these tests.

Summary of Practice Survey

Neuropsychologists in North America and Australasia were presented with a case vignette of a client with physical and sensory disabilities, and asked how they would assess such an individual. This was compared with their regular assessment practice. Regardless of their theoretical orientation, respondents generally reported making substantial changes to their standard assessment procedures. Practitioners were also canvassed regarding the domains that in their view constituted a comprehensive neuropsychological examination. The study raised issues regarding the most appropriate manner in which to adapt neuropsychological assessment measures in order to be suitable for people with expressive disabilities.

CHAPTER 6: SELECTION OF RESEARCH PROTOCOL MEASURES

Introduction

This chapter outlines the selection of the measures included in the research protocol. Two groups of measures were included. The first group of measures (the 'research measures') were selected or adapted to be suitable for use with people with expressive disabilities, and were administered to all participants. The second group of measures (the 'comparison measures') were included for domains where the research measures had been adapted from their standard administrations. The comparison measures were administered in standard format in order to investigate the effects, if any, of the adaptations made to the research measures. As these comparison measures were administered in standard format, they could only be administered to the non-disabled participants.

The present chapter commences with a description of the process by which research and comparison measures were selected, and ends with the list of tests included in the research protocol. Initial information about adaptations made to these measures is discussed, with full details reported in the following chapter, *Methodology*. Firstly, however, the present chapter begins by defining the criteria on which measures were selected.

Criteria for Suitability of Measures

In order to be considered for inclusion as a research measure[9] in the protocol, tests had to meet the following minimum criteria:

- *Response Compatibility:* The measure had to require a response which was compatible with the abilities of the disabled participant group; that is, did not require writing, drawing, a verbal response, or fine motor manipulations, such as constructing an item using test materials.
- *Timed Measures:* Response latency in individuals with severe physical disabilities may be partially due to physical rather than cognitive factors.

---

[9] The term *research measure* here, and for the remainder of this dissertation, identifies measures selected (and adapted) to be suitable for individuals with expressive disabilities, as distinct from the *comparison measures* which the two non-disabled samples also completed.

Therefore, measures were selected where performance was based solely on the content of a response, and not on its speed.

- *Fatigue:* Surprisingly little research has examined the effects of fatigue on the neuropsychological testing performance of individuals with traumatic brain injury. However, fatigue is a common sequalae of brain injury (Lezak, 1978, 1995; Montgomery, 1995) and it is possible that individuals with severe disabilities in addition to traumatic brain injuries may experience even greater fatigue from the testing process than their non-disabled counterparts. For this reason, and to ensure the assessment process was as minimally taxing for participants as possible, measures were selected which provided the most information about cognitive functioning for the least fatigue on the part of the participant.
- *Accessibility:* Although some very specialised tests have been developed for use with specific populations, few if any services have the resources available to purchase every published test, and many tests developed in research are never published. Accordingly, where possible the current project utilised adaptations of measures in common use (e.g., the Wechsler scales; Wechsler, 1997a), or measures which neuropsychologists could obtain for no cost. Additionally, where a selection of measures was possible, tests that were already available at the research site were selected over measures that would have to be purchased specifically for the research programme.
- *Psychometric Properties:* The measure had to have acceptable psychometric properties. That is, empirical research should have provided support for the validity of the measure. Additionally, research should have demonstrated the reliability of the test is consistent with theory on the stability of the underlying constructs. While robust psychometric properties were important for a test to be included, this criterion was evaluated last as, regardless of the psychometric properties, a test would not be useful if participants were unable to fulfil the task requirements. An overview of the psychometric properties of tests that were selected is presented in *Measures,* Chapter 7.

Criteria for Adaptations to Measures

Where measures did not meet all of the criteria listed above, possible adaptations to the measures were considered. To be included in the research protocol these adapted measures not only had to meet the criteria stated above,[10] but also had to meet the following additional criteria:

- *Face Validity:* Measures should continue to measure a similar construct to the original measure. While this was to be empirically examined in the research, the initial criterion was that face validity of the measure should be maintained.
- *Measure Length:* The length of time required to administer a measure should not be greatly increased. In some cases, where responses to the adapted measure might be slower than in the standard administration (e.g., where previously verbal responses were made via an alphabet board), alterations to the content of the measure, such as removing some items, were considered.

Using these criteria for adaptations, changes were proposed to the methods of administering measures that would otherwise be unsuitable. Where the standard administration already constrained responses, cards were produced from which participants could select their answer, using any reliable method[11]. Where more complex verbal responses were normally required, administration was adapted to allow participants to use an alphabet board. On the basis of anecdotal clinical experience, it was known that some individuals would be able to communicate relatively quickly by pointing to an alphabet board, while others, who might be directing an assistant with eye movements in order to use an alphabet board, could take considerably longer. Theoretically, then, any verbal response could be spelt out on an alphabet board, but some longer measures would be unsuitable. For example, Logical Memory (WMS-III; Wechsler, 1997a) could be adapted so that responses were made with an alphabet board. However, it could take some individuals as long as 30 minutes to spell out each story. Below average scores, therefore, would have to be interpreted with caution as the task of recall over this extended period is

---

[10] With the exception that psychometric properties of these newly adapted measures clearly could not have yet been established.

[11] See Chapter 7 for the operational definition of *reliable responding* that was employed in this research.

substantially more difficult than the standard administration. Due to the length of time and effort required, it was unlikely that such an adaptation would prove useful.

Sources of Potential Measures

One aim of this research project was that the research should be clinically relevant (see *Aims and Objectives*, Chapter 2). Therefore, the starting point in selecting measures to be included in the research protocol was to examine the measures used for clinical neuropsychological assessment of individuals with traumatic brain injuries at the Psychology Clinic, Massey University. As non-standard assessment is the focus of this research, measures used at this Clinic in cases where the standard measures may be unsuitable were also examined. All these measures are presented in Table 8.

In addition to the Psychology Clinic measures, alternatives were also considered and evaluated using the criteria outlined above. Firstly, tests were drawn from the two major reference compendiums of neuropsychological measures, Lezak (1995) and Spreen and Strauss (1998). Measures commonly used in research were identified from the previous review of neuropsychological research (Journal Review, Chapter 4), an examination of the protocols used by two large multicenter studies of traumatic brain injury (Kreutzer, Gordon, Rosenthal, & Marwitz, 1993; Levin et al., 1990), and an examination of research investigating the neuropsychological assessment of individuals with disabilities (reviewed in Chapter 1). Finally, research examining neuropsychological test usage among clinicians was examined (Butler, Retzlaff, & Vanderploeg, 1991; Sellers & Nadler, 1992) and neuropsychologists' assessment practices with disabled individuals were directly surveyed (Practitioner Survey, Chapter 5). These sources thus encompass current clinical and research practice, in both general neuropsychological assessment and assessment of individuals with disabilities.

Table 8. Suitability of Psychology Clinic assessment measures for individuals with expressive disabilities, blocks to their use, and possible adaptations.

Measure	Suitability	Blocks	Adaptations
Orientation			
Mini-Mental Status Examination: Orientation questions	✧	V	Alphabet board.
Attention			
Sustained Attention			
Letter Cancellation Task	✗	T, W	
Divided/Selective Attention			
Stroop Color-Word Test	✗	T, V	
Learning and Memory			
Short Term Memory Span			
Digits Forwards (WAIS-III)	✧	V	Alphabet board.
Spatial Span Forwards (WMS-III)	✗	M	
Working Memory			
Digits Backwards (WAIS-III)	✧	V	Alphabet board.
Letter-Number Sequencing (WAIS-III)	✧	V	Alphabet board.
Arithmetic (WAIS-III)	✧	V	Alphabet board.
Spatial Span Backwards (WMS-III)	✗	M	
Paced Auditory Serial Addition Test	✗	T, V	
Auditory Consonant Trigrams	✗	V	
Verbal Memory			
Rey Auditory Verbal Learning Test, or California Verbal Learning Test, or Word Lists (WMS-III)	✗	V	
Logical Memory (WMS-III)	✧	V	Multiple-choice questions only, Story A immediately, Story B at 30 minutes.
Verbal Paired Associates (WMS-III)	✧	V	Select from list of 8 responses, repeat 3 times, then at 30 minutes, followed by recognition using yes/no response card.
Nonverbal Memory			
Rey-Osterreith Complex Figure Test	✗	W	
Milner Maze	✗	M	
Digit Symbol (WAIS-III): Incidental Learning and Free Recall.	✗	W	
Faces (WMS-III)	✧	V	Yes/no response card.
Family Pictures (WMS-III)	✧	V	Alphabet board.
Benton Visual Retention Test	✧	W	Use recognition format.
Visuoperception and Visual Reasoning			
Matrix Reasoning (WAIS-III)	✓		
Picture Completion (WAIS-III)	✧	V	Alphabet board.
Picture Arrangement (WAIS-III)	✗	M, T	
Hooper Visual Organization Test	✧	V	Alphabet board.

(table continues)

Measure	Suitability	Blocks	Adaptations
Construction			
Block Design (WAIS-III)	✗	M, T	
Rey-Osterreith Complex Figure Test: Copy trial	✗	W	
Draw a bicycle	✗	W	
Draw a house-tree-person	✗	W	
Draw a clock	✗	W	
Language Functioning and Verbal Reasoning			
Comprehension (WAIS-III)	✧	V	Alphabet board.
Similarities (WAIS-III)	✧	V	Alphabet board.
Vocabulary (WAIS-III)	✧	V	Alphabet board, simplified responses acceptable, then WAIS-R-NI type multiple-choice.
Boston Naming Test	✧	V	Alphabet board.
Peabody Picture Vocabulary Test-Revised	✓		
Familiar and Novel Language Comprehension Test	✓		
Executive Functions			
Wisconsin Card Sorting Test	✧	M	Participant makes selection but assistant manipulates cards.
Twenty Questions Test	✧	V	Alphabet board.
Controlled Oral Word Association Test	✗	T, V	
Tower of London, or variant	✗	M	
Category Fluency	✗	T, V	
Information Processing Speed			
Digit-Symbol Coding (WAIS-III)	✗	T, W	
Paced Auditory Serial Addition Test	✗	T, V	
Stroop Color-Word Test	✗	T, V	
Symbol Search (WAIS-III)	✗	T, W	
Trail Making Test A, B and C	✗	T, W	
Motor Functioning			
Finger Tapping Test	✗	M, T	
Symbol Copy (WAIS-III)	✗	M, T	
Grooved Pegboard, or	✗	M, T	
Purdue Pegboard	✗	M, T	
Premorbid Functioning			
National Adult Reading Test-2	✗	V	
Personal and Emotional Functioning †			
Symptom Checklist 90 Revised	✧	W	For SCL-90-R, BDI-II and STAI, questions read out, responses on multiple-choice stimulus card.
Beck Depression Inventory-II	✧	W	
State-Trait Anxiety Inventory	✧	W	
Minnesota Multiphasic Personality Inventory-II	✗	W	

(table continues)

Measure	Suitability	Blocks	Adaptations
Functional Abilities			
Patient Competency Rating Scale	✗	W	[Item content unsuitable].
Neuropsychology Behavior and Affect Profile	✗	W	[Item content unsuitable].

Note:
Suitability: ✓ = Measure suitable for individuals with expressive disabilities; ✧ = Measure suitable with adaptations; ✗ = Measure unsuitable with individuals with expressive disabilities.

Blocks to use: M = Requires fine motor/constructional response; T = Measure depends on timed response; V = Requires verbal response; W = Requires written/drawn response.

† Administration of measures of personal and emotional functioning can be adapted for individuals with expressive disabilities, but some item content will be inappropriate. For example, measures of mood often incorporate questions regarding physical agitation, or psychomotor retardation. Responses such as "I am so restless or agitated that I have to keep moving or doing something" (BDI-II) may not be appropriate for individuals with severe motor disabilities.

The alternative measures identified from these sources are presented in Table 9. Due to the sheer number of tests reviewed in Lezak (1995) and Spreen and Strauss (1998), measures from these texts were only presented in the table if they were considered to be suitable for use with individuals with expressive disabilities (with adaptations if required).

Furthermore, measures were only included if they appeared to be as suitable as, or more suitable than, the other measures under consideration. Similarly, as there were a plethora of measures identified in the Journal Review, only measures employed by more than one study were included from this source. The two reviews of measures used by clinicians were included to identify measures likely to be available to clinicians; therefore, only measures endorsed by more than 10% of the respondents in these samples were included. Measures that were designed exclusively for the elderly or children were excluded. Finally, measures were excluded which were previously part of standard neuropsychological batteries, but which have been removed in subsequent revisions (e.g., Figural Memory from WMS-R; Wechsler, 1987). Table 9 only lists measures that had not already been considered as part of the Psychology Clinic standard measures (i.e., Table 8).

Table 9. Alternative assessment measures, blocks to use with and adaptations for individuals with expressive disabilities.

Measure	Source	Suitability	Blocks	Adaptations
Orientation				
Benton Temporal Orientation	C	✧	V	Alphabet board.
Galverston Orientation and Amnesia Test	MC	✧	V	Alphabet board.
Good Samaritan Hospital Orientation Test	L	✓		10-item form with yes/no responses.
Attention				
Sustained Attention				
Attentional Capacity Test	L	✓		
Continuous Performance Test	C, J	✗	M, T	
Digit Vigilance	J	✗	M, T	
H-R Seashore Rhythm Test	C, J, P	✦	V	Like/unlike response card.
H-R Speech Sounds Perception Test	C, J	✧	T, W	Multiple-choice, delay presentation of next item until response made.
Divided/Selective Attention				
Attentional Capacity Test	L	✓		
Brief Test of Attention	SS	✧	V	Alphabet board.
Double Letter Cancellation Test	MC	✗	M, T	
Learning and Memory				
Short Term Memory Span				
Point Digit Span	L	✗	M	
Serial Digit Learning	C	✧	V	Alphabet board.
Working Memory				
Nil.				
Verbal Memory				
Autobiographical Memory Interview	J	✗	V	
Babcock Story Recall	C	✗	V	
Buschke Selective Reminding	C, J, MC	✗	V	
California Verbal Learning Test: 9-word dementia version	J	✗	V	
Denman Neuropsychological Memory Scale: Verbal subtests	C	✗	V	
Story Memory Test	J	✗	V	
Randt Memory Test: Verbal subtests	C	✗	V	
Recognition Memory Test: Words	L, SS	✓		
Non-verbal Memory				
Continuous Recognition Memory Test	L	✧	V	Old/new response card.
Continuous Visual Memory Test	C	✧	V	Old/new response card.
Denman Neuropsychological Memory Scale: Nonverbal subtests	C	✗	V, W	
Figure Memory Test	J	✗	W	
H-R Tactual Performance Test	C	✗	M	
Memory for Designs	C	✗	W	

(table continues)

Measure	Source	Suitability	Blocks	Adaptations
Non-verbal Memory (contd.)				
Presidents Test	C	✧	M or V	Adaptation to NZ Prime Ministers, alphabet board for verbal portions, use of assistant to manipulate cards for sequencing.
Recognition Memory Test: Faces	J	✓		
Rivermead Behavioural Memory Test: Facial Recognition	L, SS	✧	V	Old/new response card.
Rivermead Behavioural Memory Test: Picture Recognition	L, SS	✧	V	Old/new response card.
Visuoperception and Visual Reasoning				
Block Pattern Analysis Test	L	✓		
Culture Fair Intelligence Test	D	✓		[Only 4 'culture fair' subtests].
Embedded Figures	C	✗	W	
Facial Recognition Test	C, D, P	✓		
Judgement of Line Orientation Task	C, J, P	✧	V	Alphabet board.
Minnesota Paper Form Board Test	L	✧	T, W	Selection via any method, remove time limit
Randt Memory Test: Picture Recognition	C	✧	V	Alphabet board.
Raven's Colored Progressive Matrices	J	✓		
Raven's Progressive Matrices	D, P	✓		
Right-Left Orientation Test	C	✧	M, T	Culver form with left/right response card; remove timing.
Test of Visual Neglect	C	✗	M	
Test of Visual-Perceptual Skills	SS	✓		
Visual Form Discrimination	C, L, MC	✓		
Construction				
Bender Visual Motor Gestalt	C	✗	W	
Boston Parietal Lobe Battery Constructional Apraxia	C	✗	M, W	
Developmental Test of Visual Motor Integration	C, J	✗	W	
Draw-a-Daisy	C	✗	W	
Three-Dimensional Block Construction	C	✗	M	
Language Functioning and Verbal Reasoning				
Aphasia Language Performance Scales	C	✗	V, W	
Apraxia Exam	C	✗	M	
Boston Diagnostic Aphasia Exam	C, J	✧	M, V, W	Word Recognition, Word-Picture Matching, and first part of Complex Ideational Material suitable; response cards, selection via any method.
H-R Aphasia Screening Test	C	✗	V	
H-R Speech Sounds Perception	C, J	✧	T, W	Multiple-choice, delay presentation of next item until response made.
Illinois Test of Psycholinguistic Abilities: Auditory Reception	L	✧	V or M	Yes/no response card.

(table continues)

Measure	Source	Suitability	Blocks	Adaptations
Language Functioning and Verbal Reasoning (contd.)				
Multilingual Aphasia Exam (MAE): Token Test	C, MC	✗	M	
MAE: Visual Naming	C, MC	✦	V	Alphabet board.
National Adult Reading Test	J	✗	V	
Neurosensory Center Exam for Aphasia	C, J	✗	M, V	
Peabody Individual Achievement Test-Revised (PIAT-R): Reading Comprehension	D	✓		
PIAT-R: Reading Recognition	D	✧	M, V	Selection via any method, verbal part adapted to alphabet board.
PIAT-R: Spelling	D	✓		
Porch Index of Communicative Ability	C	✗	M, V, W	
Token Test	J	✗	M	
Western Aphasia Battery	C	✗	M, V, W	
Wide Range Achievement Test-3: Reading	C, J	✗	V	
Wide Range Achievement Test-3: Spelling	C	✧	W	Alphabet board.
Woodcock-Johnson Tests of Achievement: Reading	J	✗	V	
Executive Functions				
Behavioral Assessment of the Dysexecutive System: Rule Shift Cards Test	SS	✧	T, V	Yes/no response card, remove timing.
Cognitive Estimation Test	L, SS	✧	W	Alphabet board.
Color Form Sorting Test	C	✗	M	
Design Fluency Test	J	✗	W	
Graphical Sequence Test	J	✗	W	
Modified Card Sorting	MC	✧	M	Participant makes selection but assistant manipulates cards.
Porteus Maze	C, D	✗	W	
Proverbs Test	C	✧	V	Use multiple-choice form.
Ruff Figural Fluency Test	C	✗	W	
Short Category Test	D, J	✓		[Use booklet version].
Tinkertoy Test	C	✗	M	
Information Processing Speed				
Symbol Digit Modalities Test: Oral	MC	✗	T, V	
Symbol Digit Modalities Test: Written	J, MC	✗	M, T	
Motor Ability				
H-R Grip Strength	J	✗	M	
H-R Lateral Dominance	C	✗	V	[Item content unsuitable for people with physical disabilities].
Sensory Functioning Measures				
Finger Localization	C	✧	V	Benton multiple-choice adaptation; requires intact tactile sensation.

(table continues)

Measure	Source	Suitability	Blocks	Adaptations
Sensory Functioning Measures (contd.)				
H-R Fingertip Number-Writing Perception	J	✧	V	Response card or alphabet board; requires intact tactile sensation.
H-R Sensory-Perceptual Examination	C, P	✧	V	Response card; requires intact tactile sensation.
Tactile Form Perception	C	✗	M	

Note:

Tests and Source: C = Tests used by clinicians (Butler, Retzlaff, & Vanderploeg, 1991; Sellers & Nadler, 1992); D = Review of disability assessment literature (Chapter 1); H-R = Halstead Reitan; J = Journal Review (Chapter 4); L = Lezak (1995); MC = Multicenter traumatic brain injury studies (Kreutzer et al., 1993; Levin et al., 1990); P = Practitioner Survey (Chapter 5); SS = Spreen and Strauss (1998). Source only identified as Lezak or Spreen and Strauss if the measure was not included from any other source; many of the other measures are also included in these volumes.

Suitability: ✓ = Measure suitable with individuals with expressive disabilities; ✧ = Measure suitable with adaptations; ✗ = Measure unsuitable with individuals with expressive disabilities.

Blocks to use: M = Requires fine motor/constructional response; T = Measure depends on timed response; V = Requires verbal response; W = Requires written/drawn response.

† Administration adaptations indicated; some item content of these measures is unsuitable for use with individuals with severe disabilities. See Note, Table 8.

Suitability of Psychology Clinic Measures

Of the 59 standard tests and subtests examined, listed in Table 8, 95% required a response that would be blocked in a client with expressive disabilities. Twenty-seven percent required a written or drawn response, 19% a fine motor/constructional response, and 41% a verbal response. Twenty-five percent of measures assessed performance at least partly on the time taken to complete test items. Based on the criteria stated above, 30 measures were excluded (labelled '✗ ' in Table 8). The criteria under which each of these measures were excluded are also displayed in the table. Only the Matrix Reasoning subtest of the WAIS-III (Wechsler, 1997a), the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981), and the Familiar and Novel Language Comprehension Test (Kempler & Van Lancker, 1996), were suitable without adaptation (5%). These measures require a simple choice between a number of stimuli, and this choice can be indicated through a number of mechanisms (e.g., pointing, verbal response).

Of the measures that required an examinee to write, draw or manipulate test materials, only one test of executive functions (Wisconsin Card Sorting Test), could be suitably adapted to another format. While standard administration of this measure requires a motor response, examinees are required only to select between alternatives, making an adaptation to a multiple-choice format potentially possible. The Benton Visual Retention Test has an alternative form, which requires a similar choice to be made. While more complex adaptations of other measures could be possible, the potential for miscommunication between examiner, participant, and any assistants, and the high risk of fatigue, made this unsuitable. An example of an adaptation considered unlikely to be successful would be the Tower of London test. While the participant could indicate which peg they wished to move a block from, and then where they wished it to be placed, there is a risk that if the assistant misinterprets an instruction, or if the participant changes their mind about a move, it may be very difficult for this to be communicated clearly. Secondly, the task would take considerably longer in this format, increasing the task requirements in terms of both cognitive demands and fatigue.

Many measures that require a verbal response could be adapted to either a forced-choice response card (e.g., yes/no), or the use of an alphabet board. However, as with the motor tasks, a number of verbal measures were excluded due to the increased complexity such an adaptation would introduce. For example, while Auditory Consonant Trigrams could theoretically be converted for alphabet board use, the process of tapping out numbers on the alphabet board was not seen as closely mirroring the original distracter task requirements. With this task, the possibility of asking participants to silently count backwards was considered also, but this was ruled out as there could be no objective analysis of whether this instruction was followed.

Fifteen of the standard measures involved timing of responses. Some of these measures (e.g., Block Design, WAIS-III) give bonus points for quick responses, and research has examined administering these tests without these time limits for disabled individuals (e.g., Berninger, Gans, St. James, & Connors, 1988). Therefore, these tests were only excluded if the responses required were incompatible despite removal of time bonuses. Berninger et al. investigated modifications to some of these measures to multiple-choice formats. However, if modification to a multiple-choice format was attempted for some measures this would change the underlying construct being assessed. For example, Block Design would become a measure of visuoperception if the constructional demands of the task were removed. Therefore, for Block Design, Berninger et al. required that their participants manipulated the blocks, but provided larger blocks that were attached to a magnetic surface. However, for individuals with no ability to manipulate test materials, this procedure is unsuitable.

For other measures, the time to complete the task is either inherent in the underlying construct, or central to the design of the measure. For example, the Stroop Color-Word Test examines the ability to shift perceptual set, follow changing task demands, and to suppress a habitual response in favour of an unusual one (Spreen & Strauss, 1998). In the Golden (1978) version of the Stroop test, examinees are required to correct any errors as they go, and to carry on with the task, with the total number of correct items in 45 seconds being the measure of performance. In this case, the time taken to complete the measure is the operational definition of the underlying construct, and as such, no time-free adaptation is possible. For this reason, measures such as this were also excluded from the protocol.

The risk of fatigue compromising performance, due to the time required for administration, led to the exclusion of the word list learning tests (AVLT, CVLT, WMS-III Word Lists). Alphabet board adaptation could have been made for the word list learning tests, but the time required for disabled participants to repeat the lists to the examiner would compromise interpretation of any difficulties as being due to a memory deficit. Also, repeating the lists a number of times with an alphabet board could lead to substantial frustration and fatigue on the part of the participants. Use of the selective-reminding procedure (Buschke & Fuld, 1974) would partially alleviate this difficulty, but would still require considerable time to administer.

Suitability of Alternative Measures

From the 95 alternative measures identified, 48 were excluded because the required response was incompatible with the abilities of individuals with expressive disabilities, and no suitable adaptations could be found. Five of these measures were also excluded because the measure required timing of responses. Adaptations were proposed for a number of measures that were otherwise not suitable. These adaptations are outlined in Table 9. Again, while many measures which usually require a verbal response could be converted to response cards or alphabet board responses, a number of measures were excluded. This was because fatigue due to the time required to administer the measure could be considerable. The measures excluded were the Autobiographical Memory Interview (Kopelman, Wilson, & Baddeley, 1989), Babcock Story Recall (Babcock, 1930; Babcock & Levy, 1940; Rapaport, Gill, & Schafer, 1968), Buschke Selective Reminding (Buschke & Fuld, 1974), CVLT 9-word dementia version (Libon et al., 1996), Denman Neuropsychological Memory Scale verbal subtests (Denman, 1984, 1987), Randt Memory Test verbal subtests (Randt & Brown, 1986), and the Story Memory Test (Heaton, Grant, & Matthews, 1991).

Selection of Research Measures

The present research project aimed to have at least one assessment measure for each domain of cognitive functioning (i.e., each domain delineated in earlier reviews; see Chapter 4 and Chapter 5). Learning and memory have been shown to be domains particularly vulnerable to disruption following traumatic brain injury, and memory contains clear theoretical subdivisions (although the practical distinctions, and their assessment, are still debated; see, for example, Heilbronner, 1992; Heilbronner, Buck, & Adams, 1989). Therefore, greater attention was given to learning and memory, with one measure selected from each sub-domain (i.e., short-term memory span, working memory, verbal memory, non-verbal memory). Other domains can be similarly refined (e.g., attention and concentration, Sbordone, 1992; executive functions, Levin et al., 1997), but there was a trade-off between having a comprehensive research protocol and practical constraints on the assessment process. In general, selecting one measure in each of the domains mirrors the starting point for clinical assessment, and provides a significantly more robust assessment than may have often been possible in the past with clients with expressive disabilities. However, where two appropriate measures assessed different aspects of a domain, inclusion of both was considered, while still balancing the need for efficient assessment.

Therefore, from the measures under consideration, which met the minimal criteria for being compatible with the communication abilities of individuals with expressive disabilities, the most appropriate measures in each domain were selected. The 68 measures from Table 8 and Table 9 considered potentially suitable (with adaptations) were evaluated, again based on the criteria described earlier. All these measures (with the suggested adaptations) could be completed by individuals with expressive disabilities. From this list, measures were selected which required the least adaptation, would minimise fatigue for participants, and, in general, were considered most readily accessible to clinicians.

Additionally, as parallel comparison measures were to be administered in their standard administration to the non-disabled groups, measures with parallel forms were favoured. It was desired that all measures selected had reasonable psychometric properties; however, due to the multiple practical constraints with individuals with expressive disabilities, the standard measures which had been researched the most and which had the most favourable psychometric properties were not always suitable.

The following sections discuss the measures that were considered for inclusion in the protocol, and outline the decisions taken for each domain. Information about the psychometric properties of the selected measures, and more detail about the administration of these measures, is presented in Chapter 7 (*Methodology*). Due to the large number of measures being evaluated, information from secondary sources that have reviewed these measures, such as Lezak (1995) and Spreen and Strauss (1998), was frequently referred to in making the comparisons in this section.

Orientation

Four measures of orientation were considered: the Mini-Mental Status Examination orientation questions (MMSE; Folstein, Folstein, & McHugh, 1975), the Benton Temporal Orientation test (Benton, Van Allen, & Fogel, 1964; Benton, Hamsher, Varney, & Spreen, 1983) the Galveston Orientation and Amnesia Test (GOAT; Levin, O'Donnell, & Grossman, 1979) and the Good Samaritan Hospital Orientation Test (Sohlberg & Mateer, 1989). These four essentially similar measures all examine orientation to time; the MMSE, GOAT and Good Samaritan Hospital Orientation Test also examine orientation to place, and the latter two of these additionally examine orientation for personal information.

The Good Samaritan Hospital Orientation Test has two forms, with the short 10-item form being specifically designed for use with individuals who cannot speak but can indicate a yes/no response. Adaptations to the other measures would either require questions to be re-written to convert them to this format, or the use of an alphabet board. The Good Samaritan Hospital Orientation Test is freely available and therefore readily accessible to clinicians. Finally, the 10-item version is as short as any of the other measures, and requires the simplest response from participants, thereby minimising fatigue. Therefore, this measure was selected for inclusion in the research protocol.

Attention

The Halstead-Reitan Seashore Rhythm Test (Reitan & Wolfson, 1989; Seashore, Lewis, & Saetveit, 1960), and the Halstead-Reitan Speech Sounds Perception Test (Reitan & Wolfson, 1993) were both considered as measures of sustained attention. The Brief Test of Attention (Schretlen, Bobholz, & Brandt, 1996; Schretlen, Brandt, & Bobholz, 1996) examines the ability to divide attention, while the Attentional Capacity Test (Webber, 1988) includes trials that measure sustained attention and selective/divided attention. These measures all require an individual to attend to an aurally presented stimulus. Both the Speech Sounds Perception Test and the Seashore Rhythm Test were originally conceived as tests of auditory perception, verbal and non-verbal, respectively. Researchers have subsequently concluded that these measures contain significant attention and concentration components (e.g., Hom & Reitan, 1990). In contrast, the Attentional Capacity Test and the Brief Test of Attention were designed specifically as measures of attention.

The Seashore Rhythm test and the Speech Sounds Perception Test are both part of the Halstead-Reitan battery and many clinicians will be familiar with their use and have access to them. The Seashore Rhythm Test requires a participant for each trial to indicate whether two musical beats are like or unlike, and can be easily adapted to response card presentation[12]. The Speech Sounds Perception Test requires a selection between four novel choices for each of the items, thus necessitating multiple response cards, but again not requiring substantial modification. An individual attends to mixed lists of numbers and letters for the Brief Test of Attention, counting first how many numbers were presented, and on a subsequent trial how many letters were presented. The required response is a single number and can be easily adapted to response card or alphabet board use. Finally, the Attentional Capacity Test consists of three trials at eight progressively more difficult levels. Each trial requires a single digit response, which can be indicated by pointing on a card, and requires no adaptation for individuals with expressive disabilities. It measures sustained and divided attention with increasingly complex tasks, providing a fine-grained picture of an individual's attentional abilities.

The Attentional Capacity Test was considered the most suitable and useful of these measures for this project. However, the shorter of the two versions of the Attentional Capacity Test still takes 25 minutes to administer, a length considered too long for the amount of information it would provide. Therefore, for the current research project a new measure, the Graded Attentional Test, was developed based on the principles of the Attentional Capacity Test. This test takes only 5 minutes to administer, and is procedurally similar to the original scale. For full details see *Measures,* Chapter 7.

---

[12] That is, a card on which each response is printed, from which participants select their response either by pointing, or directing an assistant using eye movements.

Learning and Memory

Short Term Memory Span

Two measures of short-term memory span were considered, Digits Forwards from the WAIS-III/WMS-III (Wechsler, 1997a) and Serial Digit Learning (Benton, Hamsher, Varney, & Spreen, 1983). Both of these measures require a string of digits as the response, following the aural presentation of that string of digits. Either measure could be adapted by using a number response card. Digits Forwards provides information about the limits of an individual's short-term memory span. In contrast, Serial Digit Learning does not give information on an individual's maximum digit span per se, but does provide information about learning over repeated trials. Of the measures under consideration Digits Forwards is in fact the only true measure of short-term memory span, and was selected for the research protocol. For the research measure, a new set of digit strings was generated with a random number generator[13], so that the standard strings could be used as the comparison measure (see *Selection of Comparison Measures* below).

Working Memory

Digits Backwards (WAIS-III/WMS-III; Wechsler, 1997a), Letter-Number Sequencing (WAIS-III/WMS-III; Wechsler, 1997a), and Arithmetic (WAIS-III/WMS-III; Wechsler, 1997a) were considered as measures of working memory. All three measures are drawn from the Wechsler scales, and are thus equally accessible by clinicians. All three require adaptation to alphabet board or a numeric response card.

While Arithmetic contains a significant working memory component, it also assesses an individual's arithmetic ability, confounding the two constructs. Therefore, the selection was made between the other two measures. As Letter-Number Sequencing requires a response of a string of numbers and letters, it was felt that level of

---

[13] In all cases throughout this dissertation, random numbers were generated using the RAND function in Microsoft Excel 98 for Macintosh. Technically speaking, computer generated *random* numbers are not random in the pure sense of the word, since they are usually based on a fixed formula which is seeded with a number (such as the current date and time). If seeded with the same starting value these formulae would always produce the same number sequence. However, such number sequences are pseudo-random, and can for most practical purposes be treated as random.

familiarity with the use of an alphabet board might interact with the working memory requirements of the task. It was concluded that the more simple response required for Digits Backwards would be less likely to interfere with the assessment of working memory. Therefore, this measure was selected for the research protocol. As with Digits Forwards, a new set of digit strings was created for the research measure.

Verbal Memory

Under consideration for measurement of verbal memory were three tests: Logical Memory (WMS-III; Wechsler, 1997a), Verbal Paired Associates (WMS-III; Wechsler, 1997a), and the Words subtest from the Recognition Memory Test (Warrington, 1984). Logical Memory assesses memory for information that is logically related, whereas verbal paired associates examines memory for previously unrelated word pairs, and the Words subtest from the Recognition Memory Test examines simple word recognition. As subtests of the Wechsler Memory Scale-III (Wechsler, 1997a) the first two of these measures would be more commonly available to clinicians than the latter, and have been standardised on a large sample. Secondly, the Verbal Paired Associates adapted administration incorporates more active memory requirements than the Recognition Memory Test, while maintaining a relatively simple response. Therefore, this measure was preferable over the Recognition Memory Test. As the Verbal Paired Associates and Logical Memory subtests measure distinct aspects of verbal memory, it was decided to include both in the protocol.

Non-verbal Memory

Nine measures of non-verbal memory were considered: Faces (WMS-III; Wechsler, 1997a), Family Pictures (WMS-III; Wechsler, 1997a), the Benton Visual Retention Test (Benton, 1974; Sivan, 1992), the Continuous Recognition Memory Test (Hannay & Levin, no date; Hannay, Levin, & Grossman, 1979), the Continuous Visual Memory Test (Trahan & Larrabee, 1988), the Presidents Test (Hamsher & Roberts, 1985), the Recognition Memory Test Faces subtest (Warrington, 1984), and Facial Recognition and Picture Recognition subtests from the Rivermead Behavioural Memory Test (Wilson, Cockburn, & Baddeley, 1985).

The Benton Visual Retention Test recognition format presents a visual stimulus, and then after a short delay (e.g., 10 seconds) the correct response must be selected from a 4-choice array. No standard form of this adaptation is widely available even though it has been used in some research projects. Accordingly it was not selected for the research protocol.

The Presidents Test can be administered in four forms, all of which are based around photographs of the six most recent United States Presidents, and requiring verbal naming (without photo stimuli), verbal sequencing (of their terms in office), naming from the photographs, and sequencing from the photographs. This measure could be adapted to the local political scene, but there remain drawbacks to this measure preventing its inclusion in the protocol. It must be regularly updated, and any research and norms are geographical- and cohort-specific. In addition, it is highly dependent on a person's knowledge of political events in their area. Finally, while it has high ecological face validity, it is the least pure non-verbal memory test of the nine under consideration.

Six of the seven remaining measures follow a common format: a set of target stimuli are presented, and then in a series of both targets and foils the individual being assessed must identify which stimuli are new, and which are old (i.e., identical to ones they have already seen). The measures differ, however, in the number of targets and foils, and the type of stimuli. The Rivermead Picture Recognition, Continuous Visual Memory Test, and Continuous Visual Memory Test present line drawings of various objects, animals, and plants, while the Rivermead Facial Recognition, Recognition Memory Test Faces subtest, and Faces (WMS-III) present photographs of human faces.

Finally, the Family Pictures subtest measures memory for visual information that is complex and meaningful, and is designed as a non-verbal equivalent of the Logical Memory subtest (Wechsler, 1997a). The test requires an individual to identify which members of a family were present in each picture, to identify where in the picture they were, and what they were doing. As with the verbal memory measures, the picture/face recognition measures and Family Pictures purport to examine separate aspects of non-verbal memory—isolated versus meaningful information. Therefore, it was decided to include two measures of this domain also, the Family Pictures subtest, and one of the picture/face recognition measures.

As a part of the Wechsler Memory Scale-III, Family Pictures is widely available to clinicians. In the adapted version, participants would identify who was in each picture using a response card with names of family members, and use the standard quadrant method to identify where they were, while the activities of each member would be spelt out via alphabet board. This adaptation could take longer than the standard administration. To limit fatigue, only two of the four standard pictures would therefore be used.

Photographs of people's faces date rapidly with changes in fashion such as hairstyles, and the type of glasses that people wear, as a comparison of the stimuli of the older Rivermead and more recent Wechsler subtest reveals. Therefore, as a measure backed by considerable standardisation research, commonly available to clinicians, and containing contemporary stimuli, the Faces subtest of the WMS-III was selected for inclusion in the protocol.

Visuoperception and Visual Reasoning

Eight measures of visuoperception, one of right-left orientation and five of visual reasoning were considered for inclusion in the research protocol. The measures of visuoperception were the Hooper Visual Organization Test (Hooper, 1958, 1983), the Block Pattern Analysis Test (Caplan & Caffery, 1992), Facial Recognition Test (Benton, Sivan, Hamsher, Varney, & Spreen, 1994), Judgement of Line Orientation Task (Benton, Hannay, & Varney, 1975; Benton & Tranel, 1993; Benton, Hamsher, Varney, & Spreen, 1983), Minnesota Paper Form Board Test (Likert & Quasha, 1970), Randt Memory Test Picture Recognition subtest (Randt & Brown, 1986), the Test of Visual-Perceptual Skills (Gardner, 1982), and Visual Form Discrimination (Benton et al., 1983). Matrix Reasoning and Picture Completion from the WAIS-III, the Culture Fair Intelligence Test (Cattell & Cattell, 1960), Raven's Colored Progressive Matrices (Raven, 1965) and Raven's Progressive Matrices (Raven, 1960; Raven, Court, & Raven, 1976) were considered as measures of visual reasoning. The

Culver form of the Right-Left Orientation test (Benton, 1959; Culver, 1969) was also considered for inclusion in the research protocol, but it was decided that this measure did not examine the core features to be assessed in visuoperception or visual reasoning.

From the other measures of visuoperception, the Test of Visual-Perceptual Skills was excluded because, while comprehensive, it would take considerably longer than the other measures to administer. The other measures can be divided into two groups; those which use geometric shapes as stimuli (Block Pattern Analysis Test, Judgement of Line Orientation, Minnesota Paper Form Board Test, and Visual Form Discrimination), and those which use pictures of objects or faces (Hooper Visual Organization Test, Facial Recognition Test, Randt Memory Test Picture Recognition). Of these, the latter group were favoured as providing a task closer to real-world perceptual requirements. The Hooper Visual Organization Test was selected from these because it requires more than a simple matching response, while still allowing for relatively brief (single word) answers. To shorten the length of this measure, and to provide a comparison measure, the odd-numbered items only were used as the research measure (see *Selection of Comparison Measures* below).

With the exception of Picture Completion, the tests of visual reasoning present series or matrices, and individuals being administered the test must select either which item does not belong, or the item from a selection of choices which completes the pattern. In contrast, Picture Completion requires individuals to identify what is missing in each of a series of pictures. As Picture Completion requires a more complex response, and thus would increase administration time and participant fatigue, the other scales were preferred. From these measures, Matrix Reasoning was selected as it is available to the majority of clinicians, has recent and comprehensive normative information, and could be administered without modification.

Construction

Due to the requirement of all construction measures that individuals have intact motor skills, no suitable measure of visuoconstructional abilities could be found. While adaptations to a verbal format could be imagined (e.g., directing an assistant to manipulate the materials for Block Design), these could not be realistically extended to individuals with expressive disabilities due to the time, frustration, and fatigue such adaptations would create for both examiner and examinee.

Language Functioning and Verbal Reasoning

The 16 tests considered as measures of language functioning and verbal reasoning can be divided into five groups: measures of language sounds perception, language comprehension, confrontational naming, verbal reasoning, and academic achievement in the area of language acquisition. The Halstead-Reitan Speech Sounds Perception test (Reitan & Wolfson, 1993) is a measure of how well language sounds are received and comprehended. Language comprehension is assessed by the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981), the Familiar and Novel Language Comprehension Test (Kempler & Van Lancker, 1996), the Auditory Reception subtest from the Illinois Test of Psycholinguistic Abilities (ITPA; Kirk, McCarthy, & Kirk, 1968), and the Complex Ideational Material, Word Recognition, and Word-Picture Matching subtests from the Boston Diagnostic Aphasia Exam (BDAE; Goodglass & Kaplan, 1983). The Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1978, 1983) and the Visual Naming subtest from the Multilingual Aphasia Exam (MAE; Benton, Hamsher, Rey, & Sivan, 1994) assess confrontational naming, while verbal reasoning is assessed by the Comprehension and Similarities subtests of the WAIS-III. Finally, academic achievement in the area of language acquisition is assessed by the Vocabulary subtest of the WAIS-III, the Peabody Individual Achievement Test-Revised Reading Comprehension, Reading Recognition, and Spelling subtests (Dunn & Markwardt, 1970; Markwardt, 1989), and the Wide Range Achievement Test-3 Spelling subtest (Wilkinson, 1993).

For the research protocol the areas of language comprehension and expressive language were considered the most important to focus upon. Among the language comprehension measures, the PPVT-R and the BDAE Word-Picture Matching subtest require individuals to select pictures that match with a word presented to them. The BDAE Word Recognition is the written language equivalent in that it requires matching of words spoken by the examiner to written stimuli. In contrast, the ITPA Auditory Reception and BDAE Complex Ideational Material subtests require simply yes/no responses to questions such as “Will a stone sink in water?” (BDAE). Finally, the Familiar and Novel Language Comprehension Test is similar to the Peabody Picture Vocabulary Test-Revised, but individuals must select the picture that depicts the phrase read out by the examiner; some of these phrases are concrete (‘novel’ sentences), and some metaphorical (‘familiar’ phrases), allowing for greater understanding of the nature of any language comprehension difficulties experienced. As these measures require only a forced-choice response, they all are suitable for disabled clients.

In addition to examining language comprehension itself, it was important to establish a minimum level of comprehension as the presentation of many of the other instruments in the protocol assumes intact language comprehension. Therefore, two measures were selected from this area. Firstly, the ITPA Auditory Reception questions were used to examine basic verbal comprehension and the ability to reliably make a yes/no response. This measure was selected over the BDAE Complex Ideational Material questions as there were more items (50 vs. 24), providing a clearer indication if a participant is responding in a random manner. Random responding is a consideration in any forced-choice measure, and particularly where there are only two alternatives. Secondly, the Familiar and Novel Language Comprehension Test was used to examine more complex language comprehension, including participants’ ability to understand syntactically complex and metaphorical language. This second measure was considered important as some language difficulties may not be revealed on more basic comprehension measures, but may nonetheless affect participants’ comprehension of test instructions provided for other measures in the protocol, and was selected over the other measures because of its greater complexity.

The two measures of confrontational naming are essentially the same, requiring individuals to name the objects presented in line drawings. Of these two measures, the Boston Naming Test was selected because of its wide use and general availability to clinicians. This measure is also longer (60 vs. 30 items), which allowed the measure to be divided into two while still maintaining a reasonable number of items.

Half of the items constituted the research measure, and half the comparison measure used in the two non-disabled groups (see *Selection of Comparison Measures* below).

Executive Functions

Seven measures of executive functions were considered: the Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993), the Twenty Questions Test (Laine & Butters, 1982), the Rule Shift Cards Test from the Behavioural Assessment of the Dysexecutive System (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996), the Cognitive Estimation Test (Shallice & Evans, 1978; see also Axelrod & Millis, 1994), the Short Category Test (Wetzel & Boll, 1987), the Modified Card Sorting test (Nelson, 1976), and the Proverbs Test (Gorham, 1956a, b).

Perhaps more than any other domain, the area of executive functions is the least clearly delineated, incorporating many diverse and (potentially) discrete areas of functioning. A full assessment of even a few of the specific sub-domains identified in the literature was beyond the scope of this research project. However, it was decided to include one measure of 'executive functions' in the protocol to take a first step in the assessment of these functions in individuals with expressive disabilities.

The Cognitive Estimation Test is designed as a measure of the ability to generate problem-solving strategies. The task requires estimations of various physical quantities that do not have answers that will be automatically available to respondents (e.g., "How fast do race horses gallop?"). The task requires individuals to select an appropriate response, and to evaluate its plausibility. While not widely used or commercially available, source articles give the information required to administer the measure. Some of the items in the task are geographically and cultural-specific questions (e.g., the height of the Empire State Building). Correlations with other measures of executive functioning are low, and this measure correlates more highly with measures related to general intelligence (see Spreen & Strauss, 1998). Spreen and Strauss (1998) reviewed this measure and concluded that while it had value it was "relatively new and is in need of further refinement" (p. 198).

The Twenty Questions test examines the ability to utilise efficient strategies, to perceive super-ordinate categories, and to monitor and regulate behaviour. Participants must identify which of 42 stimuli in the array has been selected by the examiner by asking questions, to which the examiner may only answer "yes" or "no". Various scoring systems have been used, with the type of questions asked by examinees (constraint, pseudo-constraint, specific hypothesis) the main variable analysed.

The Proverbs Test presents 12 proverbs and requires individuals to interpret each proverb. Standard administration calls for written responses, although there is also a multiple-choice version of the test, where four answers are provided, one of which is appropriate and abstract, and the others either concrete or common misinterpretations. There are three parallel forms of this measure.

Three of the four remaining measures being examined (WCST, Short Category Test, and the Modified Card Sorting Test) are broadly similar in administration, and examine an individual's ability to discern the rules underlying a situation. However, the WCST and the Modified Card Sorting Test have a focus on flexibility in the face of changing task demands, while, as a short form of the Halstead Category Test (Halstead, 1947; Reitan & Davison, 1974), the Short Category Test examines rule learning, and requires more complex abstraction abilities than the WCST (Bond & Buchtel, 1984). The WCST provides a measure of perseveration, and requires test takers to realise that task requirements shift periodically without warning. The Modified Card Sorting Test is an adaptation of the WCST and is a simplification in that the possibility of ambiguous feedback is removed. While all three require a forced-choice selection between four items, in their standard administration the WCST and Modified Card Sorting Test require a participant to manipulate cards rather than simply make a selection.

The BADS Rule Shift Cards Test also examines the ability to follow a rule, and to shift from one rule to another, but the rules are clearly specified, and are presented in written form on a card which is visible throughout the test administration. A yes/no response identifying whether each card matches the current rule is required.

These diverse measures examine different abilities and therefore direct comparisons are difficult. The Cognitive Estimation Test was excluded, as the measure requires further development. The Proverbs Test provides information which overlaps with the Familiar and Novel Comprehension Test and therefore was not selected as the measure of executive functioning. The Twenty Questions test was excluded, as using this measure with alphabet board responses would be considerably more time-consuming than the simple responses required for the other measures. As the two commonly used instruments in both research and clinical practice, selection was made between the Short Category Test and the WCST. These two measures were both seen to have advantages and disadvantages. The Short Category Test could be administered to individuals with expressive disabilities without modification. However, mirroring the aim in the research to only use tests that were available to clinicians, this measure was excluded, since it was not available at the research site.

Ultimately, the decision was made to trial a new measure, developed for use with people with expressive disabilities. This new measure, the Match and Shift Categories test (MASC) is broadly similar to the WCST in terms of task requirements, but only changes rules at fixed intervals that would be more obvious to examinees. Most importantly, the MASC requires only a pointing response between four alternatives. This measure is discussed in greater detail in *Measures,* Chapter 7.

Information Processing Speed

As all available measures of information processing speed were unsuitable for use with individuals with expressive disabilities, no measure of this domain could be included in the protocol.

Motor Ability

For individuals with severe physical disabilities, standard tests of motor ability clearly are inadequate, and were not included in the research protocol.

Other Measures

A number of measures were identified from the review sources that did not seem to fit into the categories above. These measures are presented Table 9 (p. 78) under the heading 'Sensory Functioning Measures'. While a number of measures were identified that could be appropriate for individuals with expressive disabilities, this was not a core area selected for assessment in this research, and accordingly these measures were not included in the research protocol.

Summary of Selection of Research Measures

The 15 research measures selected for inclusion in the protocol are presented in Table 10. They include measures of orientation, attention, short-term memory span, working memory, verbal and non-verbal memory, visuoperception and visual reasoning, language functioning, and executive functions. While for some domains no tests could be included (e.g., information processing speed), the selected measures assess the majority of the domains described in Lezak (1995).

Selection of Comparison Measures

In addition to these research measures selected or adapted to be suitable for individuals with expressive disabilities, a group of comparison measures was selected. Individuals in the two non-disabled groups included in the clinical phase of the research completed these comparison measures, in standard form, in addition to the research measures, which were administered to all groups in the same way (e.g., requiring the use of an alphabet board rather than allowing verbal responses; see Chapter 7, *Methodology*). These comparison measures were administered to allow comparison between the adapted measures and standard neuropsychological measures with established psychometric properties. Therefore, where a research measure did not require substantial adaptation from standard administration, comparison measures were not deemed required. Where comparison measures were required, however, the criteria for selection were the following:

- *Comparability:* It was intended that comparison measures should assess the same construct as their paired research measure. Therefore, comparison measures should be as similar as possible to the research measure. Ideally, an alternate form of the research measure, administered in the standard format, would be used. In some cases, a measure could be divided in half, with odd-numbered items constituting the research measure (and adapted as required) and

Table 10. Research measures included in protocol.

Measure	Adaptations
Orientation	
Good Samaritan Hospital Orientation Test	10-item form with yes/no responses.
Attention	
Graded Attentional Test	None required.
Learning and Memory	
Short Term Memory Span	
Digits Forwards (WAIS-III)	Number board; new number strings generated.
Working Memory	
Digits Backwards (WAIS-III)	Number board; new number strings generated.
Verbal Memory	
Logical Memory (WMS-III)	Yes/no questions only; Story A questions immediately, Story B at 30 minutes.
Verbal Paired Associates (WMS-III)	Select from list of 8 responses, repeat 3 times, then at 30 minutes, followed by recognition using yes/no response card.
Nonverbal Memory	
Faces (WMS-III)	Yes/no response card.
Family Pictures (WMS-III)	Names of family response card, alphabet board where required; pictures 1 and 3.
Visuoperception and Visual Reasoning	
Matrix Reasoning (WAIS-III)	None required.
Hooper Visual Organization Test	Alphabet board; odd-numbered items.
Construction	
No suitable measures.	
Language Functioning	
Boston Naming Test	Alphabet board; odd-numbered items.
Familiar and Novel Language Comprehension Test	None required.
Illinois Test of Psycholinguistic Abilities: Auditory Reception	Yes/no response card.
Executive Functions	
Match and Shift Categories Test	None required.
Information Processing Speed	
No suitable measures.	
Motor Ability	
No suitable measures.	

even-numbered items being used as the comparison measure. In cases where neither of these alternatives were possible another test designed to tap the same domain was selected.

- *Length:* While fatigue is not such a problem in the non-disabled groups, these individuals were completing a longer test protocol (the research measures plus the comparison measures) than the participants with expressive disabilities. Therefore, where there were two appropriate alternatives the shorter measure was selected.

Orientation

Modifications to the Good Samaritan Hospital Orientation Test were not required. As stated above, comparison measures were included only where research measures had been adapted from their standard administration. Therefore, a comparison measure for this test was not selected.

Attention

As the Graded Attentional Test was designed to be suitable for individuals with expressive disabilities, a comparison measure was not required.

Learning and Memory

Short Term Memory Span

For the comparison measure, the Digits Forwards subtest was used in its standard administration. An alternative set of digit strings of the same length was produced using a random number generator, and these strings were used for the research measure. The standard set of digit strings was used as the comparison measure in order to allow direct comparison to normative studies.

Working Memory

As with short-term memory span, a new set of digit strings was produced for the Digits Backwards research measure, and the standard set of digits was used as the comparison measure.

Verbal Memory

As a comparison measure to the Logical Memory Test, the Babcock Story Recall Test (Babcock, 1930; Babcock & Levy, 1940; Rapaport, Gill, & Schafer, 1968) was used. As a comparison for the Verbal Paired Associates test, the Auditory Verbal Learning Test (AVLT; Rey, 1964; Taylor, 1959) was administered to the non-disabled groups.

Nonverbal Memory

As the presentation of Faces had not been substantially adapted, no comparison measure was selected for this test.

Family Pictures contains four pictures that are each presented, after which recall is assessed immediately and 25-35 minutes later. Two pictures were to be used for the research measure, and two for the comparison measure. In this case, the usual odd items/even items split would have meant that the two research measure scenes were outdoor scenes, while the comparison measure scenes were both indoor scenes. To provide more comparable material in the research compared to the comparison measures, the first and second pictures were presented as the research measure and the third and fourth as the comparison measure.

Visuoperception and Visual Reasoning

No adaptations were required to the Matrix Reasoning test, and therefore no comparison measure was selected. The Hooper Visual Organization Test contains 30 items; odd-numbered items were used as the research measure, and even-numbered items as the comparison measure.

Language Functioning

The Boston Naming Test is a 60-item test. For this measure, the odd-numbered items were used as the research measure, and the even-numbered items as the comparison measure. As the Familiar and Novel Language Comprehension Test and the ITPA Auditory Reception subtest did not require adaptation, no comparison measures were required.

Executive Functions

The Wisconsin Card Sorting Test (Heaton et al., 1993) was administered as a comparison measure to the Match and Shift Categories test.

Additional Measures

For a number of domains—Construction, Information Processing Speed, and Motor Ability—suitable measures could not be selected for the disabled group. The inclusion of measures assessing these domains in the non-disabled groups was considered. It was possible that constructional abilities, information processing speed and motor ability might be related to performance on the research measures, comparison measures, or both, so tests of these domains were included. The measures selected were Block Design from the WAIS-III for construction, Digit-Symbol Coding from the WAIS-III for information processing speed, and both the Finger Tapping Test (Halstead, 1947; Reitan & Wolfson, 1993) and Symbol Copy from the WAIS-III for motor ability. In order to allow complete analysis of performance on Digit-Symbol Coding, the Incidental Learning and Free Recall trials were also completed.

Summary of Selection of Comparison Measures

Table 11 presents the 12 comparison measures included in the research protocol. These measures assess each of the domains in which a research measure had been adapted from its standard administration, covering short-term memory span, working memory, verbal and non-verbal memory, visuoperception, language functioning, and executive functions. Tests were also included in the comparison measures of construction, information processing speed, and motor functioning.

Summary of Selection of Research Protocol Measures

Research protocol measures were selected on the basis that they required responses compatible with the abilities of individuals with expressive disabilities, did not examine timed performances, would minimise fatigue for participants, and which were commonly available to clinical neuropsychologists. Where measures were adapted from their standard administration, comparison measures were selected, which were to be administered to participants without disabilities to examine the validity of the adapted measures.

Table 11. Comparison measures included in protocol.

Measure
Orientation
None required.
Attention
None required.
Learning and Memory
Short Term Memory Span Digits Forwards (WAIS-III)
Working Memory Digits Backwards (WAIS-III)
Verbal Memory Babcock Story Recall Test Auditory Verbal Learning Test
Nonverbal Memory Family Pictures (second and fourth pictures; WMS-III)
Visuoperception and Visual Reasoning
Hooper Visual Organization Test (even-numbered items)
Construction
Block Design (WAIS-III)
Language Functioning
Boston Naming Test (even-numbered items)
Executive Functions
Wisconsin Card Sorting Test
Information Processing Speed
Digit-Symbol Coding (WAIS-III)
Motor Ability
Symbol Copy procedure (Digit-Symbol Coding, WAIS-III) Finger Tapping Test

Summary of Part II

This section began with the Clinic Files Review, in which the focus of this programme was narrowed to those individuals with expressive disabilities. The Journal Review and Practice Survey provided information useful in determining the domains of functioning which should be assessed in this programme, with the latter also providing a number of guides for how measures could be adapted. Based on the findings of the previous chapters, the current chapter outlined the selection of the research protocol measures.

The following section presents the clinical phase of this research programme, where the research protocol selected in this chapter was evaluated in both clinical and normative populations.

PART III: CLINICAL PHASE:

EXAMINATION OF ASSESSMENT FRAMEWORK

CHAPTER 7: METHODOLOGY

Overview of Methodology

This section, Part III, presents the clinical phase of the research programme, with this chapter outlining the methodology of that phase. The design of the clinical phase is first discussed, after which the participants in the study and the examiners are described. Administration procedures are outlined, and the chapter goes on to briefly evaluate the measures which were selected for inclusion protocol in terms of their psychometric properties and the findings of previous studies. This chapter concludes with an overview of the planned data analysis procedures.

Design

Background to the Design

The severity of both traumatic brain injury and expressive disabilities vary along continua. However, for the purposes of this project four groups of participants were proposed, which represented the extremities of these continua. This design is illustrated in Table 12.

Table 12. Participant groups by brain injury and disability status.

		Brain Injury Status	
		Traumatic Brain Injury	No Traumatic Brain Injury
Expressive Disability Status	Disability	Expressive Disability Group (ED)	Physical Disability Group (PD)
	No Disability	Traumatic Brain Injury Group (TBI)	Normative Group (Norm)

Individuals were only to be included in the traumatic brain injured samples if their injury was classified as severe. (This classification is discussed in detail in *Participants* below.) Similarly, participants were only to be included in the expressive disability groups if they had a severe functional impairment in their verbal and motor communication abilities. Essentially, while some verbalisations or verbal communication might be possible, these were significantly impaired, and these participants were unable to write or draw due to motor disabilities. For the majority of these individuals, a communication modality had been established though some external means, such as an alphabet board or electronic communication device.

While individuals within these groups would vary in terms of the severity of their brain injuries, and the extent of their expressive disabilities, the within-group variation could be considered small compared to the between-group variation. Therefore, when comparing these participants with individuals who have no history of traumatic brain injury, and no expressive disabilities, traumatic brain injuries and expressive disabilities could be treated as dichotomous variables.

The initial design for the clinical phase therefore was a 2 x 2 design, with traumatic brain injury and expressive disability status as the two independent variables. The analyses for the clinical phase of this project would then examine the main effects of traumatic brain injury and expressive disabilities, and determine whether there were any interactions between these factors. The appropriate statistical technique to answer this question would be multivariate analysis of variance, with traumatic brain injury and physical disability status as the independent variables, and scores on each instrument in the research protocol as dependent variables.

The Search for the PD Group

From the outset of the research, three of the participant groups were clearly identified in client populations. Individuals with expressive disabilities (ED group), although only occasionally presenting, had been the impetus for this research. People with severe traumatic brain injuries (TBI group) were commonly seen at the Psychology Clinic, Massey University. The third group was simply a general population normative group (Norm group). The individuals with physical disabilities that caused expressive disability *in the absence of* brain injury (PD group)[14] were selected to provide a comparison for the performance of individuals in the ED group.

---

[14] These abbreviations (i.e., ED, PD, TBI, and Norm) will be used throughout the remainder of this dissertation to refer to these participant groups. The abbreviation *TBI* is used only when referring to the participant group. When referring to people with traumatic brain injuries in general, this is spelled out in full.

During the preparatory process considerable discussion was undertaken regarding the PD group with a variety of academics, psychologists, medical personnel, and professionals providing high-level disability services in a number of New Zealand cities. All discussions suggested that there was no medical condition that caused the level of expressive disability of those in the ED group, without involving neuropathology of some type. The most positive direction investigated was regarding patients who had received spinal injuries, and one previous study in this area (Berninger, Gans, St. James, & Connors, 1988) had used a comparison group of individuals with spinal injuries. Contact was made with the Christchurch Spinal Injuries Unit, one of two such facilities in New Zealand, that provides hospital care and rehabilitation for people with the most severe spinal injuries. Mr. Allan Bean, a Consultant Neurosurgeon from this unit, identified two types of patients who visited this unit and who might fit the level of expressive disability required for inclusion in the PD group:

- Patients with very high-level spinal cord injury, which affects the diaphragm and who thus require artificial ventilation. The communication difficulties are generally present in the first 6 weeks after the accident, and then resolve.
- Patients who have a lower brain stem injury resulting in a 'locked-in' syndrome. Such a person would meet the expressive disabilities criteria, but the outcome is due to a neuropathological process and thus such patients would be unsuitable for the PD group.

Mr. Bean also noted that only about one person a year fitted in to these categories (Mr. A. Bean, personal communication, March 5, 1999).

In addition to the practical problems of recruiting such a sample, assessment for this research would need to be conducted within the window of time during which they were experiencing the expressive disabilities, and there were ethical concerns about asking patients to participate in a research project so soon after a major life-altering injury. This might be justifiable were the research focussed on the treatment of those individuals themselves, but in the present research they would be acting as a control group. Inclusion of this group was deemed inappropriate.

It was concluded that there was no suitable control group with expressive disabilities in the absence of brain injury, and the PD group was removed from the design. This decision meant that different statistical analysis techniques would have to be used, and research questions re-framed. Specifically, without the PD group, multivariate analysis of variance could not be used to investigate interactions between expressive disabilities and traumatic brain injuries.

Final Design

The final design for the clinical phase of the research was a three-groups design, incorporating the ED, TBI and normative groups from the previous design. Rather than examining main effects and interaction effects of traumatic brain injury and disability status on performance on the measures, analyses were planned to examine the utility of the adapted measures for the assessment of cognitive functioning. Planned analyses included an examination of the differences of performance between the groups, and the reliability and validity of the adapted measures in each group. The healthy group would provide initial normative data on the adapted measures.

Participants

There are a number of criteria commonly used to define severity of traumatic brain injury. Glasgow Coma Scale scores provide a useful index of severity, but to use these requires access to hospital records, something that was not possible in this project. The criterion selected, therefore, for the classification of severe traumatic brain injury was a post-traumatic amnesia (PTA) of greater than 24 hours (see Lezak, 1995). In many cases, however, it was difficult to reliably establish the length of PTA due to incomplete or inadequate records. Therefore, an alternative criterion, of a loss of consciousness (LOC) greater than 24 hours, was also used. This criterion is a conservative one, as a LOC greater than 24 hours could indicate a PTA of considerably longer duration.

ED Group

Sixteen individuals who had sustained severe traumatic brain injuries, and who also had severe expressive disabilities, were initially approached and invited to participate in the ED group. The initial criterion was for participants to have had their injury after their 16th birthday, before which time they had been healthy and functioning at normal levels. The decision to exclude those individuals who had experienced early brain damage was based on the fact that a different developmental course could lead to a substantially different cognitive presentation[15]. Ultimately however, two participants were included who had sustained injuries before their 16th birthday. These individuals were included in the sample on the basis that their injuries had been similar in origin to others and because individuals in the ED group are a very small proportion of the traumatically brain-injured population, and thus participants were difficult to locate. Both of these individuals were clearly able to manage the response requirements of the measures.

The majority of participants in this group were recruited through residential rehabilitation services which specialised in traumatic brain injury, and traumatic brain injury rehabilitation day programmes. One individual had been previously seen by a psychologist now at the Psychology Clinic, Massey University, and was contacted directly and invited to participate. Finally, the caregiver of one participant contacted the researcher because of information distributed though a community group involved with brain injury.

The severity of injuries sustained by all individuals in the ED group easily exceeded the criteria stated, with a number of individuals experiencing comas that lasted a number of months. To meet the criteria for inclusion in the ED group, participants also had to have expressive disabilities which meant that they could not communicate verbally, or if they were capable of some verbalisations, communication though this modality was very limited. Participants in this group in general were also unable to write, draw, or manipulate test materials, due to co-existing difficulties with fine motor control.

---

[15] However, the findings of this research would still be useful to inform the assessment of individuals who have had early infarcts and expressive disabilities, such as those with cerebral palsy who are unable to speak. While the cognitive findings will depend, as always, on the individual, the clinical assessment issues are similar across these groups.

With six of the original individuals, completion of the full research protocol was ultimately deemed inappropriate. With one participant, while it was believed by rehabilitation staff that reliable communication had previously been established, at the time of the assessment a reliable yes/no response could not be elicited. This participant was not capable of indicating informed consent and so was not included in the sample. Three participants were excluded due to an inability to reliably use the forced-choice response cards. Finally, two individuals were excluded due to extremely slowed responses and high levels of fatigue, which made completion of even a significant minority of the protocol impractical. Further consideration of the issues raised by these individuals is presented in the *Discussion*, Chapter 9.

Of the remaining 10 participants, the normal mode of communication of two participants was via a LightWriter[16], and of three was through use of an alphabet board. While other participants did not necessarily use such devices on a daily basis, all were familiar with alphabet board communication and frequently had used computer programmes that employed touch-keyboard control pads. All participants were therefore familiar with the type of responses required, and were capable of making such responses.

Due to New Zealand's no-fault universal-cover national accident insurance scheme, none of the participants in this group, or in the TBI group, were known to be engaged in litigation regarding their injuries. (One participant in the TBI group had previously been engaged in litigation due to the particular nature of his accident, but this had been resolved in his favour prior to the research being conducted.)

Final participants in this group were 7 males and 3 females, ranging in age from 17 to 43, with a mean age of 29. Seven were of European descent, two New Zealand Maori, and one identified as part-Maori and European. Formal education of participants in this group ranged from two years of secondary school to tertiary degrees. Participants had sustained their accidents between the ages of 12 and 38,

---

[16] A hand-held computerised device with a keyboard, screen, and voice synthesiser. As words are typed into the LightWriter they appear on the screen and are read out by the machine.

and were 2 to 15 years post-accident. Brain injuries were due to motor vehicle accidents (n=1), motor bike accidents (n=3), falling from a moving object (n=1), assault (n=1), being hit by a car (n=2), and being crushed (n=1), while one person had been found injured and the details of the accident were unknown. Participants in this group were drawn from four regions in the North Island of New Zealand.

TBI Group

The 17 participants in the TBI group had all sustained severe traumatic brain injuries after their 16th birthday, which resulted in PTA or LOC of greater than 24 hours. As with the ED group, many participants had accidents of considerably greater severity, and some participants were in comas for a period of months. Participants in this group were able to communicate verbally, and had relatively intact fine motor abilities, sufficient to write, draw, and manipulate test materials.

Participants in this group were recruited from past clients seen at the Psychology Clinic at Massey University, from rehabilitation day services, and from information distributed through community organisations involved in brain injury.

Participants were 16 males and 2 females, ranging in age from 25 to 56, with a mean age of 37. Thirteen were of European descent, and five people identified themselves as both New Zealand Maori and other. Formal education of participants in this group ranged from two years of secondary education to postgraduate degrees. Participants had sustained their accidents between the ages of 17 and 51, and were 1 to 31 years post-accident. (Despite the time since some of these accidents, all participants in this group had demonstrated ongoing cognitive difficulties.) Brain injuries were due to motor vehicle accidents (n=9), motor bike accidents (n=4), falling from a height (n=1), other falls (n=1), assault with an object (n=1), being hit with a cricket ball (n=1), and colliding with a stationary object (n=1). Participants in this group were drawn from five regions in the North Island of New Zealand.

The Normative Group

Participants in the healthy normative group were recruited predominantly through public advertising, including free short advertisements placed in two community newspapers over successive weeks (weekly distribution to 33,013 and 35,930 households per issue). When provided with information one of these local newspapers ran an article regarding the research, and on another occasion the other paper ran an (unsolicited) story about the project. Posters were also placed in supermarkets, swimming pools and medical centres. Some participants were recruited through a brief address to a local community organisation, and others were recruited by word of mouth. A number of participants also referred family members, flatmates and friends to participate in the research. Finally, on one known occasion a local radio station picked up the advertisement from the local newspaper and ran the advertisement on air. While systematic records were not kept on source of contact, participants were routinely asked where they had heard about the research. It appeared that the vast majority of participants were recruited through the local newspapers. The next most common source of contact was word of mouth, including participants referring others. Poster advertising and the one radio advertisement appeared to have little impact on recruitment, with no participants identifying these as their source of contact.

The healthy normative group recruited was larger than the other participant groups, as it was intended that preliminary normative information on the adapted measures would be based on this group. They consisted of 36 males and 44 females who had no history of significant neuropathology (see *Procedure* below for details of the screening of this group). Participants in this group ranged in age from 17 to 55, with a mean age of 32. Formal education of participants in this group ranged from one year of secondary education to postgraduate degrees. Eighty-six percent were of European descent, 13% New Zealand Maori or part-Maori and other, and 1% Asian. Participants in this group were drawn from three regions in the North Island of New Zealand.

Examiners

Assessments of all individuals in the ED and TBI groups were conducted by the researcher, at all sites. However, three paid research assistants assisted with data collection in the normative group. All research assistants had undertaken specific graduate-level training in neuropsychology as part of postgraduate degrees; one had completed a Masters degree in Psychology, the other two were doctoral candidates in Psychology. Two had previously worked as paid research assistants collecting normative information on neuropsychological assessment instruments in a New Zealand elderly population.

In addition to their prior study and work in the area, research assistants were thoroughly briefed about the project, and given one-on-one training by the researcher. All research assistants also took the opportunity to practice administering the research protocol measures, becoming familiar with the specific administration instructions for the adaptations. Assistants then accompanied the researcher, with the informed consent of the participant, during an assessment of an individual in the normative group. During this assessment both the researcher and research assistant completed a datasheet, and following the session these were compared item by item, which was followed by a comprehensive debrief.

The research assistant next conducted an assessment, with the researcher also completing a datasheet. Once again, a comprehensive debrief followed, and any remaining issues were addressed. Excluding the subtest administration times (as the stopwatch was accessible only to the examiner), inter-rater reliability was calculated for these sessions. Values were 97%, 99%, and 99.9%[17], indicating an excellent level of reliability. This high level of reliability, from the first administration of the protocol by the research assistants, is attributed to a number of factors, including the use of measures with objective and specific answers, the detailed training and administration instructions (see Appendix C), and the academic background of research assistants who assisted with the project. Throughout the study, close contact was maintained between the researcher and assistants, and any issues that arose from assessments were clarified.

---

[17] Values recorded by the two raters agreed on this occasion for 879 of the 880 variables.

Procedure

Counterbalancing of Measures

To control for the possible effects of order of presentation, it was desirable to counterbalance the order of presentation of measures. However, with 13 research measures and 11 comparison measures it was not possible to employ a systematic counterbalancing using every possible presentation order. One possibility would have been to employ a random order of administration. However, seven of the memory measures had delayed recall/recognition trials. If the measures presented during these delays varied between participants this could have altered performance on the delayed task. Finally, it was important that participants have similar administration orders across groups, but the ED group were completing only the research measures. Therefore, to ensure comparability these tests had to be presented before any comparison measures for the other two groups.

In consideration of all these issues, the decision was made to group the measures into blocks, and then to counterbalance the order of these blocks. Blocks of research measures would be presented in counterbalanced order first for all participants, and comparison measures would be then be presented, again counterbalanced, to the TBI and normative groups.

The standard procedures for memory measures with recall trials usually call for the delay of a flexible but relatively specific period, such as 25–35 minutes. It was anticipated that delay times would be a difficulty in this study as it was known on the basis of clinical anecdote and the earlier studies conducted in this programme (*Clinic Files Review*, and *Practitioner Survey*), that individuals with severe disabilities would take a considerably longer period of time to complete measures. It was also known on the basis of clinical experience that some individuals with traumatic brain injuries (both ED and TBI groups) might take longer than people in the normative group, particularly on tasks with the additional requirements of using an alphabet board. In clinical practice, the order of test administration can be varied to ensure the appropriate delay period is achieved. However, this would lead to the concern raised above that with differing tasks within a delay, task requirements would differ between participants. A trade-off between variable delay times, and variable content within the delay, was required. A decision was made to hold content fixed, and allow time to vary as required to complete all measures. However, as it was expected that administration time would be systematically related to participant group (as discussed above), it was considered important to at least quantify the extent of these differences. Therefore, time was recorded at the beginning of each test, and at the end of each block, to allow test administration time, and delayed recall/recognition periods, to be calculated.

Measures were divided into two groups: tests with predominantly verbal stimuli, and tests with predominantly non-verbal stimuli. It was decided that during the delay period of verbal memory tests, measures with non-verbal stimuli would be presented. Similarly, during the delay period of non-verbal memory tests, measures with verbal stimuli would be presented (see Table 13). This would allow the administration conditions of each research measure and its comparison measure to be as similar as possible.

Measures were thus allocated to blocks. Firstly, the four research measures of memory which had delayed trials were allocated: one verbal memory and one non-verbal memory measure were randomly allocated using a computerised random number generator to each of two blocks, and their order within the blocks was also randomly allocated. These blocks are presented in Table 13.

Between the initial memory trial and delayed recall/recognition trial, three other tests were administered. This occupied a sufficiently long period for relatively fast participants, while not representing an inordinately long period for those participants who took longer to complete measures. As there were not three other tests for every memory measure in the protocol, memory measures were interlinked, so that the initial trial of the second memory measure immediately preceded the delayed recall trial of the first. See Table 13 for an illustration of this.

The other two places in each delay period were filled by random allocation of non-verbal stimuli tests in the delay periods for verbal memory measures, and vice versa.

Table 13. Administration order of research protocol measures.

Research Measures		Comparison Measures	
Verbal stimuli	Non-verbal	Verbal stimuli	Non-verbal
Block A		Block C	
Logical Memory A		Babcock Story Recall Test A	
	FANL-C		Digit–Symbol Coding
	Hooper VOT		Hooper VOT: Comparison
	Faces A		Finger Tapping Test
Logical Memory B		Babcock Story Recall Test B	
Graded Attentional Test			Wisconsin Card Sorting Test
Auditory Reception			
	Faces B		
	Match and Shift Categories Test		
Block B		Block D	
	Family Pictures A		Family Pictures A: Comparison
Digits Forwards		Digits Forwards: Comparison	
Digits Backwards		Digits Backwards: Comparison	
Verbal Paired Associates A		Auditory Verbal Learning Test A	
	Family Pictures B		Family Pictures B: Comparison
	Boston Naming Test		Boston Naming Test: Comparison
	Matrix Reasoning		Block Design
Verbal Paired Associates B		Auditory Verbal Learning Test B	

Note:
VOT = Visual Organization Test
FANL-C = Familiar and Novel Language Comprehension Test

At the end of this allocation, one research measure remained outstanding, the Match and Shift Categories test, and this was randomly allocated to Block A and placed at the end of the block, after the second delayed memory trial.

Where there were direct comparison measures for tests in Blocks A and B, these were placed in parallel positions in Blocks C and D. However, a number of research measures did not have comparison measures, and the additional comparison measures (Digit–Symbol Coding, Finger Tapping Test, and Block Design) were randomly allocated to fill these spaces. As there were fewer comparison measures, these procedures resulted in Block C containing only 6 measures.

Having allocated measures to these blocks, counterbalancing of blocks was undertaken. Four administrations of the measures were used. These administration orders are presented in Figure 6. Participants in the ED group completed only Blocks A and B, and so were allocated alternately to Administrations 1 and 2, to achieve approximately equal numbers for each. As mentioned, in order to ensure the presentation conditions of the research measures was the same across all groups, comparison measures were always presented after all research measures were completed. Therefore Administrations 3 and 4 presented research measures in the same order as Administrations 1 and 2 respectively, and then followed these with the comparison measures (see Figure 6). Participants in the TBI and normative groups were allocated alternately to Administrations 3 and 4.

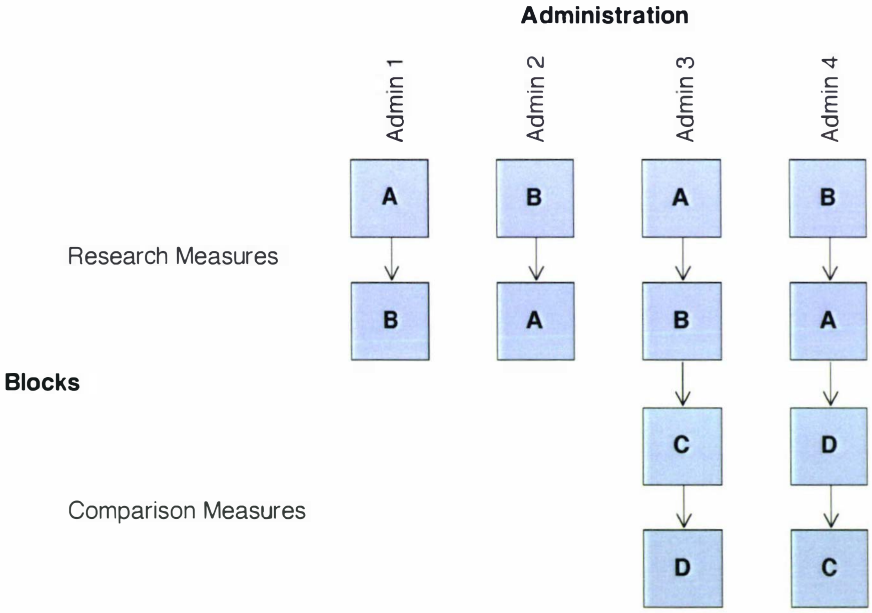


Figure 6. Counterbalancing of measure blocks in the research protocol.

Participant Screening

The screening procedure for the normative group is presented in Appendix D. Questions for individuals in this group screened for history of brain injury, stroke, heart attack, brain tumours, epilepsy, and hypoxia. Presence of any of these factors was an exclusionary criterion. Participants were also excluded if they had experienced a LOC of 5 minutes or more due to a blow to the head. A LOC of shorter duration was not included in the exclusionary criterion, due to the prevalence of very mild traumatic brain injury in the general population. While small ongoing effects have been found after mild head trauma (Binder, Rohling, & Larrabee, 1997), these studies have included people with considerably longer durations of loss of consciousness or post-traumatic amnesia than the stringent criteria used here (e.g., less than 24 hours vs. less than 5 minutes here). Furthermore, whether such effects can be causally related to mild head trauma is debatable (Binder, 1997).

Participants were also excluded if they had a history of psychiatric hospitalisation. While chronic psychiatric illness does not necessarily alter neuropsychological functioning (Palmer et al., 1997), neuropsychological status has often been found to change in many psychiatric conditions (see Sweet & Westergaard, 1997, for a review). Participants were also excluded if they had a history of heavy or chronic drug use. Additional screening questions about MRI or CT scans, and concussions without unconsciousness were included in case these elicited information about neuropathology or brain injuries that would otherwise be missed. History of any of these latter factors was not, in itself, grounds for exclusion.

Initial Assessment

The administration instructions that were followed in the assessments are included as Appendix C, and an example of the datasheet used in Administration 3 is presented in Appendix E. The Administration 4 datasheet differed from this only in the order in which measures were presented; Administrations 1 and 2 were similar, but did not include the comparison measures.

At the beginning of the assessment session a stopwatch was started. Before the start of each test, and at the end of each block, time since the assessment began was recorded from this stopwatch on to the datasheet. Participants were offered the opportunity to take breaks between blocks and usually only stopped at these times. Where other interruptions in the administration were experienced, these were likewise recorded on the datasheet. Participants were offered the choice of completing the assessment on one occasion, or over a number of sessions. Typically, individuals in the normative group completed the assessment either in one session of around 3–3$^1/_2$ hours, or in two halves. Individuals in the TBI group often divided the assessment over more than one session, and sometimes as many as four sessions, completing one block on each occasion. Although participants in the ED group completed only the first two blocks, their assessments took about as long as participants in the other groups (see *Results*, Chapter 8, for data on this issue). While the administration was maintained in the planned format as much as possible, in many cases the administration order of measures had to be altered for the ED group to allow more frequent breaks. For instance, if the administration of the two blocks needed to be split into four parts, the interlocking delayed memory tasks would be separated to allow both the initial and delayed trials of memory measures to be presented on the same day. While experimental rigor was maintained as much as possible, it was recognised from the outset that flexibility would be required with the assessment of participants with severe disabilities.

Informed Consent

An information sheet was provided, and all details of the study clearly explained, following which informed consent was obtained from participants. Any questions participants had were answered before they completed a consent form. In cases where participants were unable to complete a consent form themselves, an authorised caregiver witnessed their consent, and completed the form on their behalf. The full consent procedures are outlined in Appendix C and information sheets and consent forms are presented in Appendix F.

Response Reliability

After agreeing to participate, response reliability was formally established to demonstrate that the participant was capable of making the responses required by the measures[18]. Participants were shown seven response cards, and were asked to point to items that the assessor named. The first four were response cards used in measures in the protocol. The final three examined discrimination of number, colour, and shape, and were included to establish whether participants could perceive the

---

[18] In cases where there was a question about the ability of a participant to make a reliable response, of necessity this procedure preceded the informed consent process.

categories underlying the Match and Shift Categories test, and the Wisconsin Card Sorting Test.

Originally, it was intended that 100% performance on this measure would be required for inclusion in the research. However, this represents a criterion significantly in excess of that which demonstrates reliable performance above chance level, and individuals with severe brain injuries should not be excluded on the basis of making only one or two mistakes, even on an apparently simple task. Therefore, providing performance was above chance level, participants were not excluded.[19]

Demographics

Demographic information was collected regarding participants' traumatic brain injury history (if applicable), their ethnicity, age, highest level of formal education, and whether they were colour-blind.

Medical Screen

All participants were asked a series of medical screening questions, which covered the same areas assessed in the telephone screening of the normative group. This screening was conducted to ensure participants met the requirements of inclusion in their group. See Appendix D for details of these questions.

Administration of Measures

The Good Samaritan Hospital Orientation Test was administered next to all participants. This measure, and Response Reliability, were administered at the beginning of each assessment to ensure that all participants were sufficiently oriented and able to make reliable responses on the other measures in the protocol.

---

[19] Response cards used in establishing response reliability varied in the number of choices available to respondents, and thus the probability of making a correct response by chance varied from card to card. Therefore, calculating the lowest score that is significantly different from chance performance would be mathematically complex. However, chance performance would produce a total score of around 5 out of 28, whereas in the present study two participants scored 23 and 27, respectively, with the remainder of the sample making no errors at all. This clearly indicates performance above chance levels.

Following response reliability assessment, the research and comparison measures were administered according to the administration procedures outlined below in *Measures*, and in Appendix C. The tests were administered in the counterbalanced orders described above.

For each test, the examiner recorded participant responses on the datasheet. In many tests participants' actual responses were recorded on the datasheet. In other tests the examiner recorded whether particular answers were correct or incorrect. Generally, examiners were not required to make evaluative judgements about the scoring of tests during an assessment session. However, in tests using the alphabet board, in addition to correctly spelled words, phonetic and pseudo-phonetic responses were acceptable, providing such an answer could reasonably be pronounced in the same way as the correct answer. Where a research assistant was conducting an assessment, in any case where an answer was questionable, the participant response was recorded verbatim, and the researcher judged the accuracy of the answer at a later date.

At the conclusion of the initial assessment, participants were given $15 (see *Participant Expense Reimbursements* below), and completed the *Initial Assessment Payment and Contact* form (see Appendix F). On this form they certified that they had received the $15, and were asked whether they would be interested in participating in a follow-up assessment. Finally, participants were given the opportunity of receiving a feedback letter of the results of the overall project, to be mailed out at the conclusion of the project.

Follow-up Assessment

Follow-up assessments were conducted to examine the test-retest reliability of the research measures. There is no fixed rule for the length of retest intervals when examining test-retest reliability. It has been observed that test-retest reliability coefficients tend to be inversely related to the length of the retest interval (McCaffrey & Westervelt, 1995). The selection of an interval must depend on a theory about the stability of the underlying construct and consider the relative problems of practice effects (that presumably diminish with time) versus non-test-related change in the individuals being assessed (that presumably increase with time).

For the current research, the decision was taken to conduct follow-up assessments no sooner than 14 days, and if possible within 21 days of the initial assessment. Follow-up assessments were not conducted later than 28 days after the initial assessment. If a participant could not be seen within this time, they were excluded from the follow-up assessment sample. This retest interval was brief enough to control for major changes in the health status of disabled participants, a significant concern when working with severely disabled populations. It was recognised that, with a relatively short retest interval, practice effects could be large. Nonetheless, this would not necessarily affect the results of this research. The rationale for this argument is presented with the results of test-retest analyses, in the following chapters.

As follow-up assessments were conducted to examine test-retest reliability of the research measures, only those measures were re-administered. Participants received measures in the same administration order at follow-up as they did at initial assessment. Therefore, participants in the TBI and normative groups who were given Administration 3 received Administration 1 at follow-up, and participants given Administration 4 received Administration 2.

Even with the reduced number of measures at follow-up for the TBI and normative groups, the research protocol still took a considerable time to administer with all groups. Therefore, a smaller sample was considered adequate to examine the issue of test-retest reliability. In general, all participants were invited to participate in the follow-up assessment, but it was planned to have at least 50% of the initial sample from each group complete follow-up assessments.

A small group of participants were, however, not invited to participate in follow-up assessments. Some individuals were assessed at distant locations or out of the city that the examiner was based in, and it was not considered the best use of time and resources to conduct a follow-up assessment. The follow-up assessments of some participants fell due during public holiday periods when assessments were not being conducted. Finally, a small number of participants in the normative group chose not to complete all initial assessment measures, or behaved in inappropriate ways towards research assistants; these individuals were not invited to participate in follow-up assessments.

With the exception of the omitted tests, the follow-up assessments were conducted with the same administration procedures as for the initial assessments.

Participant Expense Reimbursements

In partial recognition of the considerable time and effort involved in the assessments (over 3 hours of testing time for the initial assessment, and $1^1/_2$–3 hours for the follow-up assessment depending on the group), and participants' expenses travelling to the assessment sites, each individual received $15 for each of the initial and follow-up assessments.

In the unusual occurrence where participants completed only part of an assessment they still received the full $15. For example, the two individuals excluded from the ED group due to extreme fatigue still received the $15 in recognition of the significant effort they had expended in completing a small part of the protocol.

Measures

Selection of the measures for the research protocol was described in Chapter 6, and the measures are presented in Table 10 (p. 99) and Table 11 (p. 103). The following section briefly describes the psychometric properties of the measures that were included in the protocol. As many of these were adapted for use in the research, the psychometric properties reported in the literature cannot be assumed to hold. However, these previous studies provide a baseline against which to compare the adapted instruments. Reliability data for the measures in the protocol are summarised in Table 14.

Table 14. Reliability of measures included in research protocol.

Measure	References
Good Samaritan Hospital Orientation Test	
There are no known studies which have included either the version of this measure included in this study, or the longer version. Therefore, the psychometric properties of this measure are unknown.	
Graded Attentional Test	
As the Graded Attentional Test was a newly developed measure, no existing reliability data were available.	
Digits Forwards and Backwards	
Internal consistency reliability coefficient = .90; coefficients for individual age groups ranged from .84 to .93[a].	Wechsler (1997a)[b]
Average generalisability coefficient = .86; coefficients for the two age ranges reported were .87 and .84[c].	
Test-retest stability coefficient = .83; coefficients for individual age groups ranged from .73 to .89[d].	
Test-retest reliability .73[e] (115 healthy adults 17 to 82 years old ; mean retest interval 20.94 days).	Youngjohn, Larrabee, & Crook (1992)[b]
Verbal Paired Associates (VPA)	
Internal consistency reliability coefficient = .93 for VPA-I, .83 for VPA-II; coefficients for individual age groups ranged from .91 to .96 for VPA-I, .73 to .88 for VPA-II[a].	Wechsler (1997a)[b]
Test-retest stability coefficient = .82 for VPA-I, .78 for VPA-II; coefficients for the two age ranges reported were.81 and .83 for VPA-I, .77 and .79 for VPA-II[c,f].	
Auditory Verbal Learning Test	
Equivalent-form retest reliability of Form 1, measured at 6 months (n=322[g]), 1, 2, and 3 years (n=47 at each period[g]) = .30 to .54 for Trial 1, .48 to .60 for Trial 5, .57 to .64 for Immediate Recall[h], .57 to .78 for Delayed Recall[h], and .28 to .64 for Delayed Recognition.	Uchiyama et al. (1995)[b]
Equivalent-form retest reliability of Form 2, measured at 1 year = .52 for Trial 1, .60 for Trial 5, .66 for Immediate Recall, .81 for Delayed Recall, and .39 for Delayed Recognition (n=37[g]).	
Alternate-form reliability = .36 for Trial 1, .49 for Trial 5, .57 for Trial 7, .60 for Trial 8, .35 for Trial 9, and .29 for Intrusions (6 month retest interval; n varies from 659 to 890[g] for these analyses).	
Reliability of total number of words learned over five trials, r= .77, post-interference performance = .70 (alternate form; retest interval = 6-14 days; 51 healthy Australian adults aged 20-67 years).	Geffen, Butterworth, & Geffen (1994)

(table continues)

Measure	References
Logical Memory (LM)	
Internal consistency reliability coefficient = .88 for LM-I, .79 for LM-II; coefficients for individual age groups ranged from .81 to .91 for LM-I, .71 to .87 for LM-II[a].	Wechsler (1997a)[b]
Average generalisability coefficient = .77 for LM-I, .79 FOR LM-II; coefficients for the two age ranges reported were .75 and .78 for LM-I, .78 and .79 for LM-II[c].	
Test-retest stability coefficient = .77 for LM-I, .76 for LM-II; coefficients for the two age ranges reported were.74 and .80 for LM-I, both groups .76 for LM-II[c,f].	
Babcock Story Recall Test	
There are only three known studies which have included this measure, and none of these have reported its psychometric properties.	
Family Pictures (FP)	
Internal consistency reliability coefficient = .81 for FP-I, .84 for FP-II; coefficients for individual age groups ranged from .73 to .88 for FP-I, .72 to .91 for FP-II[a].	Wechsler (1997a)[b]
Test-retest stability coefficient = .66 for FP-I, .71 for FP-II; coefficients for the two age ranges reported were.63 and .68 for FP-I, .68 and .73 for LM-II[c,f].	
Faces (F)	
Internal consistency reliability coefficient = .74 for F-I and F-II; coefficients for individual age groups ranged from .65 to .80 for F-I, .66 to .83 for F-II[a].	Wechsler (1997a)[b]
Test-retest stability coefficient = .67 for F-I, .62 for F-II; coefficients for the two age ranges reported were .70 and .63 for F-I, .63 and .61 for F-II[c,f].	
Hooper Visual Organization Test	
Internal consistency reliability coefficients = .82 (166 healthy college students) and .71 (74 ‘psychoneurotic’ inpatients).	Hooper (1958)[b]
Split half reliability = .80 (16 individuals with organic brain damage, 19 with ‘functional disorders’, and 33 healthy controls; age range 19-62 years; corrected with Spearman-Brown formula).	Gerson (1974)[b]
Test-retest coefficient of concordance = .86 (initial administration repeated after 6 and 12 months).	Lezak (1982, cited in Lezak, 1995)[b]
Test-retest reliability coefficient = .68 (elderly controls; 1 year interval).	Levin, Llabre, & Reisman (1991)[b]
Block Design	
Internal consistency reliability coefficient = .86; coefficients for individual age groups ranged from .76 to .90[a].	Wechsler (1997a)[b]
Test-retest stability coefficient = .82; coefficients for individual age groups ranged from .80 to .88[d].	

(table continues)

Measure	References
Matrix Reasoning	
Internal consistency reliability coefficient = .90; coefficients for individual age groups ranged from .84 to .94[a].	Wechsler (1997a)[b]
Test-retest stability coefficient = .77; coefficients for individual age groups ranged from .75 to .81[d].	
Auditory Reception	
Internal consistency reliability coefficients ranged from .90 to .96 across age groups (962 healthy children aged 2 yrs 7 mths to 10 yrs 1 mth; scores corrected for restricted intelligence range).	Paraskevopoulos & Kirk (1969)[b]
Test-retest stability coefficients over 5 month retest interval ranged from .63 to .79 (198 healthy children, aged 4 yrs, 6 yrs, and 8 yrs; scores corrected for restricted intelligence range).	
Boston Naming Test	
Test-retest correlations over three annual administrations ranged from .62 to .89 (122 nerologically intact, independent and high functioning 57 to 85 year olds).	Mitrushina & Satz (1995)[b]
Familiar and Novel Language Comprehension Test (FANL-C)	
No studies have reported reliability data for this measure to date.	
Match and Shift Categories test (MASC)	
As the MASC was a newly developed measure, no existing reliability or validity data were available.	
Wisconsin Card Sorting Test	
Generalisability coefficients: Total number of errors, .71; Perseverative responses, .53; Perseverative errors, .52; Non-perseverative errors, .72 (normative study of 46 children and adolescents; test manual does not report any reliability studies with adult populations).	Heaton, Chelune, Talley, Kay, & Curtiss (1993)[b]
Digit-Symbol Coding	
Test-retest stability coefficient = .86; coefficients for individual age groups ranged from .81 to .91[d].	Wechsler (1997a)[b]
Test-retest reliability = .88[e] (115 healthy adults aged 17 to 82 years old; mean retest interval 20.94 days).	Youngjohn, Larrabee, & Crook (1992)[b]
Finger Tapping Test	
Test-retest reliability over 6 month period = .71 for dominant hand, .76 for non-dominant hand (healthy volunteers aged 16-70 years; n > 120[i]).	Ruff & Parker (1993)[b]
Test-retest reliability = .80 for dominant hand, .82 for non-dominant hand (60 healthy college students; 1 week retest interval).	Morrison, Gregory, & Paul (1979)[b]

Note:

[a] Normative study of 2450 American adults 16-89 years.

[b] Sample collected in the United States of America.

[c] Normative study of 297 American adults 16-89 years.

(table continues)

---

[d] Normative study of 394 American adults 16-89 years; average retest period 34.6 days.
[e] Subtest used was from original WAIS (Wechsler, 1955).
[f] Average retest period 35.6 days.
[g] Homosexual and bisexual HIV-seronegative males.
[h] No data for six month retest.
[i] Authors do not report the exact sample size for this analysis.

In addition to the information presented here, the reader is also referred to the sources of these measures, and texts such as Lezak (1995), Mitrushina, Boone, and D'Elia (1999), and Spreen and Strauss (1998), for a more detailed description of the psychometric properties of the scales. A wider literature review was conducted but due to the breadth of the literature and the number of measures under review, considerable use was made of the review sources listed above in the selection of key studies. Their contribution to the following sections is acknowledged.

Each measure is described below, and a brief summary of key studies examining the validity of the measures is presented. Particular focus was placed on studies that included samples with traumatic brain injuries, as this was the population of the current research. Many studies have investigated the properties of most of these measures, and it is possible to provide only a brief outline here. As a result some areas which are of importance to clinical practice, such as the effects of age and education on test performance, are not discussed, although research has indicated that these are important factors for some of these measures. However, since the majority of analyses conducted in this study were within-participant analyses (e.g., correlations between different test scores; see *Statistical Analyses* at the end of this chapter), these issues were of less relevance.

Each description below concludes with an outline of the changes in format of stimulus presentation and responses required for the measures adapted for this research. As a result of these changes, alterations were also required to the instructions given to participants regarding the task requirements. The administration instructions for all the tests included in the research protocol are presented in Appendix C. Additions to standard administration instructions have been underlined in that Appendix. In some cases portions of the original instructions were also omitted in making these alterations. Therefore, for a comprehensive evaluation of changes to administration instructions, the original test manual should be consulted.

Orientation and Attention

Good Samaritan Hospital Orientation Test

The Good Samaritan Hospital Orientation Test (GSHOT; Sohlberg & Mateer, 1989) is a 20-item verbally administered brief test of orientation to personal information, place, time and general information, and was developed for the longitudinal study of post-traumatic amnesia. For the purposes of the current study the 10-item short version of this measure was used, which requires only a yes/no response and which examines personal information (six items), orientation to place (two items) and orientation to time (two items). The questions from the measure are presented in Appendix E.

Sohlberg and Mateer (1989) do not report psychometric properties for this measure, and to date neither version of the GSHOT (long or short form) has been used in any known published studies. Therefore the psychometric properties of the measure are unknown.

As this measure did not require adaptation, no comparison measure was included in the protocol.

Graded Attentional Test (GAT)

The Graded Attentional Test was a new measure created for this research, based on the general format of the Attentional Capacity Test (see *Selection of Research Protocol Measures,* Chapter 6, for a discussion). The GAT is postulated to examine sustained attention. A stimulus card is displayed, which remains visible throughout the trial, indicating the number an examinee should attend to for that trial (the stimulus cards for this test are presented in Appendix G). Then, a series of numbers is read to the examinee, who must count the number of times the target stimulus is included in the list. As the test aims to examine sustained attention (and therefore lapses in attention to the stimulus list) rather than other functions such as working memory, it was considered appropriate to allow participants to manage the additional task requirements as they saw fit. Therefore, while examinees were not encouraged to use any strategies for keeping track of the number of presentations of the target stimulus, they were not prevented from doing so either. For example, participants might count on their fingers the number of presentations. It was recognised that it would be futile to attempt to eliminate this behaviour since such activities can be carried out covertly. Additionally, the instruction not to use such strategies might in fact prompt their use.

The target stimuli and lists of numbers used in the GAT are presented in Appendix C. All numbers were selected using a computerised random number generator. Firstly, initial target stimuli (4, 9, 6, 2) were selected randomly. These were combined in various forms for the target stimuli in each trial (e.g., '4 or 9' in Trial 3; '4 followed immediately by 9' in Trial 5). Secondly, the number of times the target stimulus was presented in the list was generated, constrained to a number between 1 and 10.

Trial 1 consisted of a list solely of the target stimulus (4) repeated the number of times selected above. For other trials, number lists were generated, with list lengths selected in order to allow sufficient space for the necessary number of target stimuli. These comprised 15 digits for trials 2–4, and 20 digits for trials 5–7 (as these later trials had a two digit target stimulus). Finally, randomly generated lists were modified where required to add the previously selected number of target stimuli to the lists. In Trials 5–7, lists were also modified to ensure that distracter stimuli, similar to the target stimuli, were included in the lists. These included reverse presentation of the target stimuli (e.g., 9–4 instead of 4–9) and one target stimulus without its partner. These were included to ensure that, for the later trials which were intended to be more difficult, a simple count of one part of the target stimulus would not result in the correct answer. Finally, early informal pre-testing suggested that this measure could represent a significant challenge for some participants. Therefore, a discontinue criterion of three errors in a row was set.

As this was a new measure, created for the current study, no research has previously been conducted on its reliability and validity. Further, as it was not an adapted version of a standard measure, no comparison measure was included in the protocol.

Short Term Memory Span, and Working Memory

Digits Forwards and Backwards (WAIS-III/WMS-III)

Digits Forwards (Wechsler, 1997a) is a measure of short term memory span for numbers, while Digits Backwards has been described as a complex span task (Wechsler, 1997a) or a working memory task (Lezak, 1995). Digits Forwards requires an individual to repeat (in the order they were read out) lists of numbers of increasing length. Digits Backwards requires the repetition of similar lists, but repeated in the reverse order to which they were read out.

In the Wechsler scales, Digits Forwards and Backwards are summed together for scoring, with the authors arguing that the tasks ultimately examine a single dimension. Other authors (e.g., Lezak, 1995) argue strongly that the two tasks are disparate and should be treated as such. In the present research, scores from these two tests were analysed and discussed separately. However, the standardisation studies only report reliability data on this subtest for the combined Digits Forwards and Backwards score, and it is these that are discussed here.

Some studies have reported reduced Digit Span scores following traumatic brain injury, with some studies finding greater deficits in Digits Forwards (e.g., Black & Strub, 1978) while others relate more severe lesions to a reduction in Digits Backwards (Leininger, Gramling, Farrell, Kreutzer, & Peck, 1990). However, a reduction in Digit Span is not an inevitable result of brain injury (Black, 1986).

The Digits Forwards and Digits Backwards tests were modified for this research so that, rather than repeating numbers verbally, participants gave their answers by selecting them from a number board (see Appendix G for the board used). As the numbers were being read out, however, the number board was turned upside down or covered. This was done to limit the extent to which participants could substitute or partially utilise a visual span in generating their responses.

With the exception of one trial on Digits Forwards (Item 8, Trial 1), the WAIS-III/WMS-III Digits Forward and Digits Backward do not repeat digits within a trial. Therefore, for the research measure, strings of numbers between 1 and 9 were generated using a computerised random number generator, with the provision that no number be repeated within a string. Number of trials and string length were identical to those found in the WAIS-III/WMS-III Digits Forward and Backward. The strings used in the research measure are presented in Appendix E.

For the comparison measure the original Digits Forward and Digits Backward subtests were used, administered in their standard format.

Verbal Memory

Verbal Paired Associates (WMS-III)

The Verbal Paired Associates test is a measure of cued verbal memory and learning. In the standard administration, examinees are read eight word pairs, after which the first word from each pair is presented verbally, and participants must name the word with which it was previously paired. This procedure is repeated four times. After a delay a further recall trial is completed, without the list being read again. Finally, the WMS-III provides for a yes/no recognition trial, where 24 word pairs are read, and examinees must identify whether each pair was one of the word pairs presented earlier.

The test is sensitive to left hemisphere lesions (Chelune & Bornstein, 1988), which is the location of language centres in the majority of the population. This finding may therefore indicate the measure taps verbal memory abilities. However, the words in this list can be easily visualised, and some factor analyses of the WMS-R failed to hold up the verbal versus non-verbal factor distinctions of the scales as a whole (Leonberger, Nicks, Goldfader, & Munz, 1991; Leonberger, Nicks, Larrabee, & Goldfader, 1992). However, the verbal versus non-verbal memory factor structure was upheld in the standardisation study of the WMS-III (Wechsler, 1997a). Verbal Paired Associates has been shown to be sensitive to a range of neurological disorders, such as Alzheimer's disease (Kaszniak, 1986) and in mild, apparently asymptomatic, traumatic brain injury (Stuss et al., 1985).

For the current research, the format of this measure was altered so that, following reading of the word pairs, participants were presented with a multiple-choice response card (see Appendix G). For each word, participants pointed to the partner word on the card. As with the standard administration, the list of word pairs was read four times. During the reading of each list, the response card was turned upside down.

During standard administration of this test, examinees hear the correct answer to every question; either they hear themselves say the correct answer, or following an incorrect answer the correct answer is given to them verbally by the examiner. To make the adapted version of the test as similar to the original as possible, the feedback provided by the examiner was modified. The correct answer was not being *spoken* by participants, so the examiner provided verbal repetition of the correct answer following both correct and incorrect answers. (In the standard administration of the measure only an indication that the answer was correct is provided if the examinee provides a correct answer.) See the administration procedures in Appendix C for the specific wording of this feedback.

This measure had been substantially adapted for the research, so another measure of verbal memory and learning, the Auditory Verbal Learning Test, was included as a comparison measure.

Auditory Verbal Learning Test (AVLT)

The Auditory Verbal Learning Test is a test of verbal memory and learning over repeated trials. A number of different administrations of this test have been used by various researchers; the following is one common variant, and represents the administration used in the current project. Other versions are outlined in Lezak (1995), Spreen and Strauss (1998), and Mitrushina, Boone, and D'Elia (1999).

A 15-item list of unrelated one- and two-syllable words are presented at one second intervals, after which examinees recall as many words as possible from the list. No points are awarded for the order in which items are recalled. Following recall the procedure is repeated a further four times; each time examinees are asked to recall all words, including those they recalled on previous trials. An alternative 15-item list is then presented, followed by recall for this new list. Following this, recall is tested for the first list, without repetition of the list by the examiner. After a 25–35 minute delay, recall is again tested without re-presentation of the list. Finally, recognition of words from the first and second lists is tested by having examinees identify the words from a 50-word array which includes all of the words from both lists and words which are related to the lists, either semantically, phonetically, or both.

The AVLT has been shown to tap similar functions to other measures of verbal learning (e.g., CVLT; Crossen & Wiens, 1994). It is sensitive to the effects of traumatic brain injury, with examinees typically performing below average on the free recall trials, but at normal levels on the recognition trials, suggesting individuals with traumatic brain injury are experiencing retrieval problems (see Bigler, Rosa, Schultz, Hall, & Harris, 1989; O'Donnell, Radtke, Leicht, & Caesar, 1988). However, in a similar measure, the CVLT, other authors have suggested free recall–recognition discrepancies do *not* indicate retrieval deficits (Wilde, Boake, & Sherer, 1995). Performance on the AVLT is altered in a range of neuropathological conditions, with the pattern of reduced scores differing across conditions (see Lezak, 1995, for a discussion).

The AVLT was included in the protocol as a comparison measure for the adapted version of the Verbal Paired Associates test.

Logical Memory

The Logical Memory subtest of the WMS-III examines verbal memory for logically-related information. Examinees are read a 25-unit story, after which immediate recall is examined by the participant repeating as much of the story as can be remembered. Next, a second 25-unit story is read, with a similar recall trial. This second story is read once more, and a second immediate recall trial is conducted. After 25–35 minutes the examinee is asked to retell both stories; the stories are not re-read. The WMS-III also provides yes/no recognition questions for each story.

Due to the relatively recent change in the administration of this measure with the introduction of the WMS-III, the majority of studies investigating this test were conducted with earlier versions. Individuals with traumatic brain injuries have been found to have significantly lower scores on this measure than matched trauma controls (Dikmen, Machamer, Winn, & Temkin, 1995). In addition to traumatic brain injury the test has been shown to be sensitive to a range of neuropathologies (e.g., dementia, Storandt, Botwinick, & Denziger, 1986; Wilson & Kaszniak, 1986; and multiple sclerosis, Minden, Moes, Orav, Kaplan, & Reich, 1990). Some indices from the measure have been shown to be sensitive to left versus right hemisphere infarcts, providing support for the assertion that the measure examines verbal as opposed to non-verbal memory and learning (e.g., Chelune & Bornstein, 1988).

For the current research, the administration of this measure was altered significantly. Rather than using a free-recall procedure, participants were asked only to answer the yes/no recognition questions, and to do so by pointing to their answer on the Yes/No response card (see Appendix G). However, were the immediate recognition questions re-used for delayed recognition, the answers given might represent recall of the previous answers, rather than recall of story material. Therefore, participants were read Story A, after which the recognition questions were asked. Then, Story B was read to participants. Recognition questions for this story were presented after a delay. Immediate recognition of story detail was examined with Story A only, and delayed recognition with Story B only.

The Babcock Story Recall Test was administered as a comparison measure to the adapted Logical Memory Test.

Babcock Story Recall Test

The Babcock Story Recall Test uses a 21-unit story to test immediate recall, and was administered in the format described in Lezak (1995). The story is read, and examinees are then asked to repeat as much of the story as they can recall. Immediately following this, the story is re-read. After approximately 20 minutes delayed recall is assessed. As the story is re-read the test also involves a learning component. See Appendix E for the Babcock story.

The format of this test is similar to a number of story recall tests. However, beyond the initial studies (Babcock, 1930; Babcock & Levy, 1940) only four studies that have used the Babcock Story Recall Test could be located, one which was a normative study (Kreutzer et al., 1986), while the other three used the test with clinical populations (Gabrys, 1979; Kupke, Lewis, & Rennick, 1979; Rapaport, Gill, & Schafer, 1968). Reliability and validity data for the scale have not been reported, and despite being similar to other story recall measures commonly used with individuals with traumatic brain injury, no studies using the test have previously been conducted in this population.

As this test has been infrequently used, and well-researched and standardised scoring criteria have not been reported, additional comment is required for this measure. Scoring was based partly on the principles outlined in Rapaport, Gill, and Schafer (1968). However, that source penalised participants for the intrusion of distorted or bizarre information into the story (that is, points would be subtracted from the total score). The merits of this system of scoring could be argued; it does not, however, provide a close analogue to the scoring used in the Logical Memory test. Therefore, to provide a better comparison to that measure for the current study, participants were given credit for correct information, but were not penalised for intrusions. In line with the principles of Rapaport et al., marks were allocated for correct meaning, rather than the specific words used by participants. The scoring criteria used for this test are presented in Appendix H. As there was a degree of interpretation required, participant responses were recorded verbatim by examiners, and scoring for all participants was completed at a later date by the researcher to ensure consistency.

Non-verbal Memory

Family Pictures (WMS-III)

The Family Pictures subtest of the WMS-III is designed to be a non-verbal analogue to the Logical Memory subtest, in that it assesses memory for non-verbal logically-related information. Standard administration involves the introduction of the family members and the dog in a family picture, followed by presentation of four scenes for 10 seconds each. After this, examinees are asked to recall who was in the scenes, where they were (selected from four quadrants presented on a response card), and what they were doing. Delayed recall, 25–35 minutes later, repeats these same questions for each scene.

The Family Pictures subtest represents an entirely new test, not based on any measure previously used in clinical practice or research (Wechsler, 1997a). Validity studies published from the development of the scale examine only the indexes of the WMS-III, and do not report findings of individual subtests. Furthermore, no studies could be located to date which have reported the use of this measure in any clinical or other population. Research establishing the validity of this measure is clearly required.

In the present research, a presentation similar to the standard one was used, except that when asked who was in each scene, family members' names were selected from a multiple-choice response card. As the position of family members is already selected from such a card in the standard administration, this was used. Finally, participants were asked to spell out on an alphabet board what each family member was doing in the scene[20].

In Chapter 6, it was anticipated that adaptation to this format would considerably lengthen the administration time of this test, and it was decided to use only two of the four scenes for the research measure. This provided the opportunity to use the other two scenes, in standard administration, as the comparison measure. As only two scenes were being used in this comparison measure, the task difficulty would presumably be reduced. However, the memory requirements would be similar to

---

[20] Individuals in the TBI and normative groups were always required to use the responses described. However, in this test, and with all measures, individuals in the group with severe expressive disabilities used their normal mode of communication. In some cases, this involved using a LightWriter, and in one case communication was conducted using eye movements to direct assistants with an alphabet board. In any case where a caregiver was assisting with an assessment, it was ensured that they could not see the stimulus materials. This was important to maintain the integrity of the research, but also allowed caregivers to employ strategies that they would usually use when communicating with the participant. For instance, not being aware of the correct answer, caregivers were free to guess the final word when it was partially completed, and look for confirmation from participants, a technique commonly used to speed communication. Qualitative observation of this process indicated that this was a useful technique, with caregivers occasionally confused or surprised regarding the words participants were spelling out, which were, however, correct responses for the measures.

those in the research measure, and thus this comparison measure provided the opportunity to examine the effect of response card and alphabet board use on test performance.

In other measures in the protocol that were divided in half to create the research and comparison measures, items were split odd-even. However, in Family Pictures the pictures alternate between indoor and outdoor scenes. Therefore, an odd-even split would have resulted in one measure containing two indoor scenes, and one measure with two outdoor scenes. For task comparability it was considered more appropriate to include one scene from each setting, in each measure. Therefore the first two scenes were used in the research measure, and the final two scenes in the comparison measure.

Faces (WMS-III)

The Faces subtest was included in the latest WMS-III in recognition of the findings from other researchers that tasks involving memory for faces are sensitive to right hemisphere, right temporal lobe, and hippocampal lesions (see Wechsler, 1997a). A series of 24 individual colour photographs of faces are presented at 2 second intervals, after which the examinee is presented with 48 further faces one at a time, and asked to indicate whether each face is one presented earlier. The standard administration calls for a verbal yes/no response. After a 25–35 minute delay, a further group of 48 faces are presented, and examinees are again asked to indicate whether each face was one that was presented in the first group.

Like the Family Pictures subtest, Faces is new to the WMS-III. However, similar measures such as the Warrington Faces Test (Warrington, 1984) have been widely used in clinical research and practice. Such measures have been found to be sensitive to right hemisphere lesions (Carlesimo & Caltagirone, 1995; Schweinberger, Buse, Freeman, Schönle, & Sommer, 1992), and right temporal lobe and hippocampal lesions (Wechsler, 1997a). However, while similar measures have proven useful in the past, only one study could be located which has examined the properties of the WMS-III Faces subtest. Millis, Malina, Bowers, and Ricker (1999) examined the factor structure of the WMS-III, and concluded that the Faces subtest loading on the visual memory construct was low enough to make it questionable whether the subtest should be included in this factor. In light of this finding, research is required to clarify the properties of this measure.

Faces requires only a yes/no response, so the only adaptation made was for participants to give their response by pointing to yes or no on a response card (see Appendix G). This adaptation was considered relatively minor, and no ideal comparison measure could be identified, so no comparison measure for Faces was included in the protocol.

Visuoperception and Visual Reasoning

Hooper Visual Organization Test (HVOT)

The Hooper Visual Organization Test consists of 30 line drawings of basic objects, which have been cut up into several parts and rearranged on the page. The examinee's task is to name the object, thus requiring a mental rebuilding of the presented 'jigsaw'.

There has been considerable debate as to what the HVOT actually measures. Hooper (1958) designed the test simply to identify organic brain pathology, and to differentiate such pathology from "functional and motivational disorders" (p. 1). The revised edition manual argued that the test was not specific to "visuopractic functions" but rather sensitive to generalised impairment (Hooper, 1983, p. 6). Some authors have suggested the test taps a general visual-spatial intelligence factor (e.g., Johnstone & Wilhelm, 1997). Other authors have challenged this argument, although studies which link performance to particular areas of the cortex have reported some conflicting findings (e.g., Fitz, Conrad, Hom, Sarff, & Majovski, 1992; Rathburn & Smith, 1982; Woodward, 1982; c.f. Lewis et al., 1997; Wang, 1977; Wetzel & Murphy, 1991; see also Boyd, 1981, 1982a, b).

Both Lezak (1995) and Spreen and Strauss (1998) state that low performance cannot be attributed solely to visuospatial functioning, and that naming abilities must be considered in performance on this measure. Other studies have found, however, that naming is not significantly related to scores on the HVOT (Paolo, Cluff, & Ryan, 1996; Ricker & Axelrod, 1995).

For the present research the HVOT was divided into halves, with odd-numbered items being used for the research measure, and even-numbered items being used for the comparison measure. For the research measure, rather than naming the objects verbally, participants spelt out their answers on an alphabet board. Even-numbered items used for the comparison measure were administered in the standard format.[21]

Matrix Reasoning (WAIS-III)

Matrix Reasoning was designed as a measure of four types of non-verbal reasoning: pattern completion, classification, analogy, and serial reasoning (Wechsler, 1997b). A matrix of stimuli are presented in which there is one missing part—examinees must select the missing part from five choices. Examinees select their choice either by pointing, or by identifying the choice by number.

The WAIS-III included this new subtest in order to answer criticism that aspects of fluid reasoning were not well assessed with the previous versions of the scale (Wechsler, 1997a). This scale follows a similar format to a number of previously published tests, such as Raven's Progressive Matrices (Raven, 1960, 1965; Raven, Court, & Raven, 1976; see Lezak, 1995 for discussion of these other measures). However, to date, as with the other new tests of the WAIS-III and WMS-III, no studies have reported the properties of this measure in a clinical sample.

The response required for this measure is flexible, requiring participants only to point, or indicate the number of their answer. This measure required no adaptation for use in the current research. Therefore, no comparison measure was included in the protocol.

---

[21] Part-way through the data collection period the administration of this measure was altered by an error in re-binding the stimuli booklet. See *Limitations of the Research,* Chapter 9, p. 214.

Construction

Block Design (WAIS-III)

This measure of constructional abilities requires participants to copy two-dimensional, two-colour, designs of increasing complexity using red and white blocks. The test is timed and bonus marks are given for speed of correct performance. Each block has two red sides, two white sides, and two sides half-red and half-white.

Reduced Block Design performance has been associated with posterior regions in the brain, and the parietal lobe is particularly implicated (e.g., Warrington, James, & Maciejewski, 1986). The measure is sensitive to a range of neuropathological conditions, but differentially so. For example, after mild to moderately severe traumatic brain injury, scores on this test are frequently amongst the highest of tests administered (Correll, Brodginski, & Rokosz, 1993), and in one large, well-controlled study of outcome one year after severe traumatic brain injury, test scores did not differ from matched controls (Levin et al., 1990). In contrast, for Alzheimer patients, scores on Block Design may be among the lowest (e.g., Larrabee, Largen, & Levin, 1985; Storandt, Botwinick, & Danziger, 1986).

Block Design was not included in the protocol as a specific comparison measure for one of the research measures, but rather to cover a domain not able to be assessed by any of the research measures. The inclusion of this test was to allow examination of the relationship of the research measures to constructional abilities. The other tests administered to the comparison groups to assess areas not examined by the research measures were Digit–Symbol Coding and the Finger Tapping Test. As with all comparison measures, these tests were administered only to the non-disabled groups.

Language Functioning

Auditory Reception (ITPA)

Auditory Reception is a test of auditory comprehension, and requires only a yes/no response to 50 sensible and nonsensical questions (e.g., 'Do dogs eat?', 'Do bananas telephone?'; see Appendix E for the full list of questions). The measure is part of the Illinois Test of Psycholinguistic Abilities, originally developed for use with children with developmental disabilities. Questions in the Auditory Reception subtest are concrete and thus suitable for individuals with limited educational backgrounds; however, items do become increasingly complex.

This measure was developed for use with children with developmental disabilities, and no studies could be located that included individuals with traumatic brain injuries. However, the measure was considered useful for the current study. It could provide evidence of whether individuals were responding at above chance levels, and thus being both capable of making reliable responses, and also of verbal comprehension. All tests in the protocol rely on verbal instructions, making verbal comprehension an important domain to quantify.

The standard administration of Auditory Reception allows examinees to provide a yes/no response through any medium. However, while no 'adaptation' to the standard administration was required, in the present research participants were required to give their response by pointing to *yes* or *no* on a response card (see Appendix G). That is, although allowed to respond verbally in the standard administration, participants in the TBI and normative groups were not permitted to respond verbally in the current research.

Despite this requirement, it was not considered that this adaptation was likely to alter the task requirements considerably, and a comparison measure for Auditory Reception was not included in the protocol.

Boston Naming Test

The Boston Naming Test is a test of confrontational naming. Examinees are asked to name line drawings of up to 60 common and increasingly less common objects. If unable to name an object after 20 seconds, examinees are given a stimulus clue (e.g., for the object *comb*, the clue is 'used for fixing hair'). If an examinee still is unable to name the object, a phonemic clue is given (e.g., for *comb*, "it starts with the sound c*o*"). Standard administration begins with Item 30, with prior items only administered if any of the next 8 items are failed. Among a number of indices, total naming score is calculated as the number of items named spontaneously, or with a stimulus cue.

The Boston Naming Test has been shown to be sensitive to word-finding difficulties after traumatic brain injury (Lezak, 1991, cited in Lezak, 1995) and to demonstrate naming difficulties in aphasic patients (Margolin, Pate, Friedrich, & Elia, 1990). The test has also been shown to be sensitive to even mild impairment in dementia (Storandt, Botwinick, & Danziger, 1986; Storandt & Hill, 1989), and sensitive to subcortical disease (e.g., multiple sclerosis; Lezak, Whitham, & Bourdette, 1990, cited in Lezak, 1995). Therefore, while highly sensitive to neurological infarcts, low performance on the measure is not specific to particular conditions.

In the current research, the Boston Naming Test was divided into two halves, with odd-items making up the research measure and even-items the comparison measure. The number of items in each scale was thus reduced, and all 30 items on each of these new scales was administered. For the research measure, participants gave their responses by spelling out the name of the object on an alphabet board.

As with all tests involving the alphabet board, participants were told that this was not a spelling test, and that all they had to do was "get across the general idea". Responses were marked as correct if participants provided a correct spelling of the word, or if they provided a phonetic or pseudo-phonetic spelling of the word (see *Procedure* above). That spelling was not being scored was particularly emphasised on this test since the other task demands were modest for many participants and it was thought the spelling aspect of the test could otherwise receive undue attention and cause anxiety.

The comparison measure was administered in the standard format, with the only exception being that all items from the first card were administered.

Familiar and Novel Language Comprehension Test (FANL-C)

The FANL-C examines comprehension of novel and familiar sentences. The test compares comprehension due to lexical analysis (postulated to be related to left hemisphere language functioning) and the understanding of meaning derived from the gestalt, in familiar phrases (related to right hemisphere language functioning), which would often have a substantially different meaning if analysed literally. The test presents 20 novel sentences (e.g., 'He's chasing after a white duck') and 20 familiar sentences (e.g., 'He's turning over a new leaf'), and examinees indicate which of four pictures best represents each sentence. See Kempler and Van Lancker (1987) for an example of the stimuli used in this test.

The FANL-C has been used in a number of studies by its authors. The measure has been found to be sensitive to left versus right hemisphere infarcts in adults, with left hemisphere damage lowering comprehension of novel sentences, while right hemisphere damage lowered performance on the familiar items (Kempler, Van Lancker, Marchman, & Bates, 1999; Van Lancker & Kempler, 1987). Interestingly, performance on familiar sentence items was found to be more markedly reduced than novel sentence comprehension in very early Alzheimer's disease, leading the authors to suggest that this test might be useful in the early detection of this condition (Kempler, Van Lancker, & Read, 1988).

No studies to date have reported data on the reliability of the scale (Daniel Kempler, personal communication, December 17, 1999). However, normative information has been collected on healthy individuals (n=335; stratified by individual years from 3 years to 19 years old, and decade groups for 20s to 80s) and four clinical populations (total n=61; people with left hemisphere damage, right hemisphere damage, Parkinson's disease, and Alzheimer's disease).

This measure only requires a pointing response to select between four alternatives, so no adaptation of this measure was required for the present research, and no comparison measure was necessary.

Executive Functions

Match and Shift Categories Test

The Match and Shift Categories test was designed for this research, along similar lines to the Wisconsin Card Sorting Test. Thirty cards were presented in a booklet, each containing four options. A stimulus card was also presented, containing between one and four copies of a shape in a particular colour. The test requires examinees to select which of the four options matches the key card, on the basis of an undisclosed rule. Responses can be given by pointing to the selected option, or indicating the letter of the option selected.

Appendix C presents the full administration instructions for this measure, but for convenience they are briefly stated here. Examinees are given feedback as to whether they have matched the cards correctly. Until three correct trials in a row are achieved, feedback is "No, that doesn't match in the way that I'm thinking of", or "Yes, that matches in the way that I'm thinking of". Following three correct trials, feedback is reduced to simply "Yes", reverting to the more detailed answers only if further errors are made. If, after 10 trials, an examinee still has not achieved three correct matches in a row, feedback is extended to, "No, that doesn't match in the way that I'm thinking of. This one matches in the way that I'm thinking of", and the examiner indicates the correct card. These additional procedures were introduced following informal pre-testing which suggested that the task format obscured the underlying requirements more in the Match and Shift Categories Test than in the Wisconsin Card Sorting test.

Unlike the Wisconsin Card Sorting Test, changes in category are not made during the 30 trials in the Match and Shift Categories Test. Following completion of the 30 trials (or after 20 trials if 3 correct responses have not been established), the examiner returns to the beginning of the booklet, and the same 30 cards are used for the next category. A third category is similarly repeated.

As the Match and Shift Categories Test was designed for the current project, no psychometric data are available for it. This measure was included as initial examination of the properties of such an adaptation, and was not considered to necessarily be a final version of the measure. A number of arbitrary decisions were made about issues such as discontinue criteria. It was anticipated that should the measure prove useful, further research would need to be conducted and considerable further adaptation might be needed.

This newly created measure was designed to assess aspects of the executive functions, and was based on similar ideas to the Wisconsin Card Sorting Test. Therefore the latter measure was included in the research protocol as a comparison measure to the Match and Shift Categories test.

Wisconsin Card Sorting Test

The purpose of the Wisconsin Card Sorting Test is to examine people's ability to perceive the rules underlying a situation, having established a cognitive set to maintain it, and the ability to alter their approach to a situation following changing task demands. Examinees are required to match up to 128 cards to four key cards, each time being given feedback on whether they are accurate in their match or not. Cards can be matched according to colour, form, or number of items on the card. Following 10 correct matches in a row, the rule is changed without any indication to the examinee. The test continues until either all 128 cards have been used, or 6 categories have been completed (i.e., 10 correct matches in a row). This test was administered in the standard format.

The Wisconsin Card Sorting Test has been widely researched, and has been found to be sensitive to impairment in a range of populations with neurological infarcts and conditions (see the test manual, Heaton, Chelune, Talley, Kay, & Curtiss, 1993, for a review). It has been referred to by some writers as a test of frontal lobe functioning, and neuroimaging studies have provided some evidence for frontal specificity (e.g., Arnett et al., 1994; Rezai et al., 1993), but the specificity of the test to frontal functioning has also been challenged (Mountain & Snow, 1993; Reitan & Wolfson, 1994b). While the frontal lobes may play a primary part in performance it has been suggested that the measure may examine complex executive functions resulting from the interaction of broad areas (Axelrod et al., 1996).

The Wisconsin Card Sorting Test was included in the research protocol as a comparison measure for the newly developed Match and Shift Categories test.

Information Processing Speed

Digit–Symbol Coding (WAIS-III)

Digit–Symbol Coding is thought to examine attention and information processing speed, and also depends on visual scanning and fine motor speed. Digit–Symbol Coding was administered as an additional measure to cover a domain of cognitive functioning which could not be examined with the research measures.

This test requires examinees to copy symbols that are paired with numbers, and is a timed measure. Next, participants are asked in the Incidental Learning procedures first to recall and reproduce the symbols that were matched with each number from 1 to 9, and following this to recall all the symbols they remember, regardless of which number they were associated with. Symbol Copy is similar to Digit–Symbol Coding but requires only the copying of symbols, and can thus be used to investigate whether fine motor speed might be responsible for low performance on the primary measure.

Digit–Symbol Coding is the test most consistently sensitive to brain injury in the WAIS, being affected regardless of the site of a lesion (Lezak, 1995). It has been found to be sensitive to even mild brain injury, and is therefore frequently included in sideline assessments of possible concussion in sports (e.g., Hinton-Bayre, Geffen, & McFarland, 1996). It is also very sensitive to a wide range of other neuropathologies, from dementia (e.g., Storandt & Hill, 1989; Larrabee, Largen, & Levin, 1985) to the effects of chronic alcoholism (e.g., Tamkin & Dolenz, 1990).

In the present research Digit–Symbol Coding was administered in the standard format, including the additional Incidental Learning and Symbol Copy procedures. Where a participant did not complete to the end of the fourth line in the two minutes allotted for Digit–Symbol Coding, their progress was noted for this scale but they were allowed to continue to the end of the fourth line, as is the standard administration when employing the Incidental Learning procedures. The only deviation from standard administration was that, while Wechsler (1997b) suggests that Symbol Copy procedures should be administered at the end of the testing session, this procedure was administered immediately following the Incidental Learning trial.

Motor Ability

Finger Tapping Test

The Finger Tapping Test is a measure of fine motor speed, and was administered to complete the range of domains that it was planned to assess with the research measures, where possible.

It has been found that performance on this measure is sensitive to brain injury (e.g., Dikmen, Machamer, Winn, & Temkin, 1995), and indicative of laterality of infarcts (e.g., Bigler & Tucker, 1981; Finlayson & Reitan, 1980; Reitan & Wolfson, 1994a, 1996), although other studies have indicated that there are a range of explanations for lateralised differences on this measure (e.g., Lewis & Kupke, 1992; Thompson, Heaton, Matthews, & Grant, 1987). Tapping frequency can be reduced following a range of disorders of motor functioning (Shimoyama, Ninchoji, & Uemura, 1990) and after traumatic brain injury (Haaland, Temkin, Randahl, & Dikmen, 1994).

Many administrations formats of the Finger Tapping Test have been employed. For this research, after an initial practice session, five trials of 10 seconds were conducted with the index finger of each hand alternately. Instructions used by Russell and Starkey (1993) were employed, with participants directed to tap as fast as possible using only the index finger, while keeping the 'heel' of the hand on the board, and avoiding using the whole wrist or arm. In the event that a participant did not follow these instructions on a trial, the score was discarded and the trial repeated.

Data Handling Procedures

All data were coded using Dataprep, a data coding application developed by Massey University Computing Services staff, and then imported into SPSS for Macintosh 6.1.1 (SPSS Inc., 1995) for analysis. A portion of the data was coded by the researcher. All data entered by the researcher was verified using the verification feature of Dataprep, which requires all data to be entered twice. Any mismatches are flagged during the second entry, and can be corrected. The majority of the data were coded by a professional data entry service run by Computing Services, Massey University. Missing data were manually checked and verified; likewise, the data set was screened for extraneous values. See *Distributions and Missing Data* in Chapter 8 for details about the extent of missing data.

Statistical Analysis Procedures

Control of Type I Error

From the measures included in the assessment protocol, a wide range of indexes and scores could have been calculated. However, it was considered important to limit the number of scores to be calculated from the dataset as, in general, for each additional variable on which an analysis is repeated, the risk of a Type I (false positive) error increases. There were a number of strategies that could have been employed to control for Type I error. Bonferroni alpha adjustments could have been used to minimise the chances of a Type I error in repeated analyses, but this 'blind' approach increases the risk of Type II (false negative) errors in proportion to the number of variables in the analyses, and is a conservative strategy. In considering such an approach the costs and benefits of Type I and Type II errors must be weighed up. For the current research project, Type II errors were considered as harmful as Type I errors, and such a conservative strategy was deemed inappropriate.

An alternative approach, utilised in this research, was to compute planned comparisons only, and use standard alpha levels. Also, where appropriate, omnibus multivariate analysis of variance (MANOVA) techniques were used to examine whether there was any effect of an independent variable on a group of dependent variables, and the effects on individual dependent variables were only examined if the omnibus approach indicated there were significant differences. In this case, Bonferroni alpha adjustments are not required as the risk of Type I errors has been controlled though the use of the initial omnibus analysis. These procedures were selected as a balance between sensitivity to hypothesised relationships, and control of Type I error inflation.

Another factor to be considered in determining the appropriate assessment strategy and the number of summary scores to calculate, is the size of the underlying dataset, and thus the number of genuine relationships which might be present in the data. In the current research, with a dataset of almost 900 variables per full assessment, the number of calculated indexes (presented below) was judged reasonable.

It has been argued by a number of authors that inferential statistics have been frequently overused and misused in the history of psychology, and that their use has in fact held back the development of parsimonious psychological theory and knowledge (e.g., see Schmidt, 1996). Meta-analysis is proving successful in demonstrating stability in areas of research where findings previously appeared to be highly divergent (Schmidt, 1996), and this method of analysis holds considerable promise. Some authors therefore argue that the sizes of effects should be the focus of research, rather than rejecting or failing to reject a null hypothesis. Such a view, held by the writer, sees individual studies as single datum in a field, rather than reporting the final answer to any research question. For this reason, rather than structuring results as a dichotomy (significant/non-significant), it is, in general, more appropriate to report the size of an observed effect and a confidence interval for this value. These effect size coefficients[22] are in a standard metric that is independent of the type of analysis or sample size, allowing comparisons between studies to be made.

There are cases where statistical significance testing is more appropriate, particularly where on the basis of the findings a two-alternative forced choice must be made. For example, when using a statistical test to determine whether a distribution is different from a normal distribution, a researcher may wish to know whether to use parametric or non-parametric statistics. In this case, the conclusion of a significant or non-significant difference from a normal distribution is more useful than a measure of effect size. Significance testing is used for this purpose in this research.

---

[22] $$ES = \frac{\bar{X}_E - \bar{X}_C}{SD_{pooled}}$$

Where
ES = Effect size.
$\bar{X}_E$ = Experimental group mean.
$\bar{X}_C$ = Control group mean.
SD_{pooled} = Pooled standard deviation.

In the following analyses, effect sizes are reported where appropriate. Inferential statistics are additionally reported, to allow comparison of results against the standard (though perhaps at times inadequate) reporting practice in the field. In the text, statistics are reported in the following format: statistic name, degrees of freedom, statistic value, p-value, and effect size. For example: F(2, 105)=13.28, $p < .001$, $f = 0.50$. Formulas used to calculate effect sizes are provided where appropriate.

Unless otherwise stated the statistical analyses were conducted using SPSS for Macintosh 6.1.1 (SPSS Inc., 1995). Chi^2 analyses reported in earlier studies, inter-rater reliability calculations, and effect size calculations (except where stated) were computed manually using Microsoft Excel 98 for Macintosh (Microsoft Corporation, 1998). Confidence intervals for correlations were computed using the calculation engine on the Vassar College web site (http://faculty.vassar.edu/%7Elowry/rho.html).

Selection of Scores for Inclusion in Analysis

The summary scores that were calculated for use in the following analyses are presented in Table 15. A number of compromises were required to reduce the number of variables included. For example, the Digit Symbol Coding incidental learning procedures were used, but to reduce the number of variables, the Incidental Learning and Free Recall scores were summed into one Digit Symbol Memory index. In other cases, where a range of scores could be calculated from a measure, only selected variables were used. For instance, information about intrusions and interference trial performance on the Auditory Verbal Learning Test was not included in the analyses. Again, a balance had to be struck between inclusion of this additional information, and the resultant increase in the probability of either Type I or Type II errors.

Table 15. Scores calculated from protocol measures.

Measures and Score Labels[a]	Calculated Score
Response Reliability	
Response Reliability	Total correct.
Good Samaritan Hospital Orientation Test	
GSH Orientation	Total correct.
Graded Attentional Test	
Graded Attentional Test	Number of trials correct.
Digits Forwards and Backwards	
Digits Forwards	Number of trials correct.
Digits Backwards	Number of trials correct.
Digits Forwards comp. [b]	Number of trials correct.
Digits Backwards comp.	Number of trials correct.
Verbal Paired Associates, and Auditory Verbal Learning Test (AVLT)	
Verbal Paired Associates A	Sum of Trials A-D.
Verbal Paired Associates B	Score on delayed recall trial.
VPA Recognition	Score on recognition trial.
AVLT Trials 1-5	Sum of all List A words correctly recalled from Trials 1-5.
AVLT Delayed Recall	Number of List A words correctly recalled after delay.
AVLT Recognition	Number of List A words correctly recognised after delay.
Logical Memory, and Babcock Story Recall	
Logical Memory A	Correct answers to Story A recognition questions.
Logical Memory B	Correct answers to Story B delayed recognition questions.
Babcock Story A	Total story items recalled on initial recall + 4 points[c].
Babcock Story B	Total story items recalled on delayed recall.
Faces	
Faces A	Total correct on initial trial.
Faces B	Total correct on delayed trial.
Family Pictures	
Family Pictures A	Total score on correct items.
Family Pictures B	Total score on correct items.
Family Pictures A comp.	Total score on correct items.
Family Pictures B comp.	Total score on correct items.
Auditory Reception	
Auditory Reception	Total correct.

(table continues)

Measures and Score Labels	Calculated Score
Boston Naming Test	
Boston Naming Test	Number of items correct, incl. correct with stimulus cues.
Boston (all cues)	Number of items correct, incl. stimulus or phonetic cues.
Boston Naming Test comp.	Number of items correct, incl. correct with stimulus cues.
Boston (all cues) comp.	Number of items correct, incl. stimulus or phonetic cues.
Familiar and Novel Language Comprehension Test (FANL-C)	
FANL-C Novel	Total correct on novel items.
FANL-C Familiar	Total correct on familiar items.
Hooper Visual Organization Test	
Hooper VOT	Total score.
Hooper VOT comp.	Total score.
Block Design	
Block Design	Total score.
Matrix Reasoning	
Matrix Reasoning	Total correct.
Match and Shift Categories Test (MASC), and Wisconsin Card Sorting Test (WCST)	
MASC Total Correct	Total correct on Items 1-20 of Form and Colour trials.
MASC Categories	Number of items till 10 correct in a row, summed for both Form and Colour trials.
MASC Perseveration	Number of items that match the previously established category, on the first 20 items of the Colour and Number trials. (Only calculated if participant has established each previous category, defined as 10 correct answers in a row).
WCST Total Correct	Total number of correct responses.
WCST Perseveration	Total number of perseverative responses.
Digit Symbol Coding	
Digit Symbol Coding	Total number of items correctly coded.
Digit Symbol Memory	Sum of Incidental Learning and Free Recall scores.
Symbol Copy	Total number of items correctly copied.
Finger Tapping Test	
Finger Tapping	Overall mean score across both hands, of 5 trials of 10-second duration with each hand.

Note:

[a] The labels listed are used to refer to these variables for the remainder of the dissertation, and these abbreviations are not defined again in future tables.

[b] Comp. = Comparison Measure.

[c] Four points were added to the initial recall score as per the procedures outlined in Lezak (1995).

Reliability

To examine reliability of the research measures, internal consistency and test-retest reliability calculations were planned for each of the measures. Internal consistency calculations were also performed on those comparison measures which were matched to specific research measures, as a baseline against which to compare the adapted measures. As the comparison measures were not included in the follow-up assessments, test-retest reliability could not be calculated for these measures.

Validity

Analysis of the intercorrelations between research measures and their comparison tests were planned. It was hypothesised that, if the adapted and new measures were valid measures of their underlying domains, they should be highly correlated with their comparison measures included in the protocol. Furthermore, it was expected that correlations between measures theorised to examine similar constructs should be stronger than correlations with other tests. However, even with the selectively reduced set of variables calculated for the analysis (see Table 15), the potential number of intercorrelations that could be examined was enormous. Calculation of this many correlation coefficients again raises questions of Type I error inflation. Therefore, as a method of examining the interrelationships of all variables in the analysis, an exploratory factor analysis was planned.

Additionally, it was hypothesised that as the majority of research measures have been found to be sensitive to the effects of traumatic brain injury (in their standard forms), scores on the adapted measures would reliably differentiate between participants in the three groups. As it was a reasonable hypothesis that the severe expressive disabilities of the ED group were associated with more severe brain injury, it was hypothesised that there would be a statistically significant difference between all three groups on these measures, with the ED group performing at a lower level than the TBI group, who would perform at a lower level than the normative group.

Normative Data

It was planned to develop preliminary normative data from the assessments of the normative group. Where possible, data from the normative group were converted into standard scores, based on a normal distribution with a mean of 10 and a standard deviation of 3. As this format is used in the Wechsler scales, this system is very familiar to clinicians. In cases where the distribution of scores on a particular scale made such a conversion questionable (e.g., the majority of individuals in the normative group scoring on or very close to the test ceiling), percentile norms were presented as well.

CHAPTER 8: RESULTS

Overview of Results

This chapter describes the results of the quantitative analyses of the clinical phase. *Qualitative* findings of this phase are presented in Chapter 9, along with the discussion of the *quantitative* analyses presented here. For the present, distributions of the variables are examined first, and the way in which outliers and missing data were handled is outlined. Then, the psychometric properties of the research measures are evaluated. Firstly, test administration and delayed recall times for the measures in the research protocol are briefly examined. Next, reliability of the research measures is evaluated in terms of internal consistency and test-retest reliability. Validity is next examined, from the perspectives of concurrent validity, construct validity, and discriminant validity. Following this, normative data for the research measures are presented. The chapter concludes with a close examination of the performance of the target group—those with expressive disabilities—to determine whether each of the adapted measures is a suitable instrument for the assessment of cognitive functioning in this group. Having presented the results of all these analyses, Chapter 9, *Discussion*, relates these findings to the research questions posed at the outset of the programme, regarding the assessment of individuals with expressive disabilities.

Distributions and Missing Data

Distributions

Means and standard deviations of research protocol measure scores are presented in Table 16, and for the associated comparison measures in Table 17. Distributions were examined to determine whether variables were normally distributed. Normal distributions were not expected in the ED and TBI groups due to small sample sizes, and the general tendency for large individual differences between people with brain injuries. Visual examination of histograms and normal probability plots of the normative group scores, however, revealed that even in this group some variables appeared to display significant skewness and kurtosis, while others displayed an exponential growth-curve-shaped distribution, with increasing frequencies to the ceiling of the measure. Such findings of non-normal distributions are not unusual in neuropsychological data, and may in fact be the norm (Retzlaff & Gibertini, 1994).

Table 16. Means and standard deviations of research measure scores by group.

Group	ED	TBI	Norm
Good Samaritan Hospital Orientation Test			
Mean	9.20	9.72	9.95**
Std Dev.	(1.03)	(0.67)	(0.22)
Graded Attentional Test			
Mean	2.30	5.67	6.26**
Std Dev.	(2.06)	(1.53)	(0.98)
Digits Forwards			
Mean	5.60	8.50	10.86
Std Dev.	(3.06)	(1.79)	(2.47)
Digits Backwards			
Mean	2.80	5.72	8.48
Std Dev.	(1.87)	(1.36)	(2.51)
Verbal Paired Associates A			
Mean	4.60	15.22	26.80**
Std Dev.	(2.46)	(7.77)	(6.14)
Verbal Paired Associates B			
Mean	1.10	4.11	7.44**
Std Dev.	(0.88)	(2.63)	(1.18)
Verbal Paired Associates Recognition			
Mean	16.70	22.83	24.00[a]
Std Dev.	(4.64)	(2.20)	—
Logical Memory A			
Mean	9.80	11.83	13.36**
Std Dev.	(1.81)	(1.98)	(1.40)
Logical Memory B			
Mean	8.90	10.06	12.50*
Std Dev.	(1.73)	(3.00)	(1.85)
Faces A			
Mean	24.40	28.11	35.96
Std Dev.	(2.67)	(4.39)	(4.93)
Faces B			
Mean	25.70	32.17	37.46
Std Dev.	(3.92)	(3.82)	(4.45)
Family Pictures A			
Mean	6.70	16.00	25.81*
Std Dev.	(2.21)	(5.09)	(4.89)
Family Pictures B			
Mean	8.60	15.56	25.60
Std Dev.	(2.50)	(5.75)	(4.96)
Auditory Reception			
Mean	41.10	48.39	48.46**
Std Dev.	(6.35)	(2.06)	(2.37)
Boston Naming Test			
Mean	17.63	25.22	27.34**
Std Dev.	(5.60)	(4.05)	(2.96)
Boston Naming Test (all cues)			
Mean	22.38	27.61	29.14**
Std Dev.	(5.53)	(2.93)	(1.96)
FANL-C Novel			
Mean	8.70	18.53	18.81**
Std Dev.	(4.57)	(1.33)	(2.08)
FANL-C Familiar			
Mean	12.60	19.29	19.69**
Std Dev.	(4.03)	(1.05)	(0.82)
Hooper Visual Organization Test			
Mean	9.63	22.61	25.81
Std Dev.	(5.80)	(3.29)	(2.88)
Matrix Reasoning			
Mean	6.10	15.78	19.29
Std Dev.	(3.90)	(4.02)	(4.05)
MASC Total Correct			
Mean	25.38	33.17	35.40**
Std Dev.	(8.78)	(5.64)	(4.22)
MASC Categories			
Mean	41.50	31.44	26.81**
Std Dev.	(13.53)	(10.53)	(6.93)
MASC Perseveration			
Mean	9.50	10.00	10.14**
Std Dev.	(3.11)	(0.85)	(0.98)

Note:

* Significant deviation from a normal distribution, p < .05 (** p < .01).

[a] All this group scored 100% on this test.

Table 17. Means and standard deviations of comparison measure scores by group.

Group	TBI	Norm
Digits Forwards comp.		
Mean	8.47	11.15
Std Dev.	1.77	2.49
Digits Backwards comp.		
Mean	5.47	7.67*
Std Dev.	1.28	2.46
AVLT Trials 1-5		
Mean	38.18	55.65
Std Dev.	10.98	9.30
AVLT Delayed Recall		
Mean	6.71	11.82
Std Dev.	4.73	2.71
AVLT Recognition		
Mean	10.29	13.73**
Std Dev.	4.03	1.53
Babcock Story A		
Mean	11.65	14.90
Std Dev.	2.89	3.48
Babcock Story B		
Mean	8.24	14.72
Std Dev.	5.92	3.92
Family Pictures A comp.		
Mean	17.53	23.18
Std Dev.	6.23	4.66
Family Pictures B comp.		
Mean	16.71	22.87
Std Dev.	5.68	4.69
Boston Naming Test comp.		
Mean	25.41	27.66**
Std Dev.	4.49	3.00
Boston (all cues) comp.		
Mean	27.59	28.91**
Std Dev.	3.12	1.95
Hooper VOT comp.		
Mean	25.00	27.62**
Std Dev.	3.10	2.27
Block Design		
Mean	37.00	50.68
Std Dev.	10.76	10.60
WCST Total Correct		
Mean	73.41	68.80**
Std Dev.	12.35	8.98
WCST Perseveration		
Mean	19.76	8.13**
Std Dev.	18.16	6.03
Digit Symbol Coding		
Mean	54.82	80.61
Std Dev.	13.15	15.92
Digit Symbol Memory		
Mean	14.24	22.24
Std Dev.	5.53	4.51
Symbol Copy		
Mean	98.24	120.90*
Std Dev.	22.96	14.66
Finger Tapping		
Mean	84.85	96.85
Std Dev.	11.10	12.56

* Significant deviation from a normal distribution, p < .05. (** p < .01)

Kolmogorov-Smirnov tests of normality calculated for all variables confirmed that in the normative group more than half of the variables in the current study differed significantly from a normal distribution. The specific variables that differed are indicated in Table 16 and Table 17.

The extent to which parametric statistics are robust to violations of the assumption of normality continues to be debated (see, for example, Grimm & Yarnold, 1995). Transformations to achieve normality were considered, but any such transformations complicate the interpretation of results (e.g., logarithmic transformations change scores to a measure of relative change, or score elasticity; Hair, Anderson, Tatham, & Black, 1998). A decision was made to use raw scores in all analyses, despite deviations from normal distributions. To guard against the possibility of misleading results due to the effects of distribution shape, non-parametric statistics were calculated in tandem with parametric analyses for group comparisons. This strategy avoids the significant loss in power of using only non-parametric statistics, while providing a balance against the risk of error. In analyses where participants acted as their own controls (e.g., test-retest reliability calculations), underlying distributions were not important, and only parametric statistics were used.

Outliers

It was expected that a number of statistical outliers would be observed in each of the samples, across a range of measures. Statistical outliers are scores which are statistically extreme compared to the rest of the sample from which they were drawn, and raise the question as to whether the individual concerned is an appropriate representative of a particular group. In this research statistical outliers were expected for two reasons, the first due to a *decrease* in score variability (due to the measures used), and the second due to an *increase* in score variability (due to participant characteristics).

Decreases in score variability occurred because a number of the tests included in the research protocol sampled only the low end of the ability range in a particular domain. The inevitable change in task requirements when multiple-choice administration formats are used generally results in a test becoming less difficult. In addition, the range was deliberately set at a relatively low level for some measures, to ensure that all participants, including those with severe cognitive impairments, would be able to manage at least parts of the tasks. This process was likely to produce statistical outliers however, since in the normative group ceiling effects would occur, meaning that the normal tail of the distribution might appear to be more disparate from other test scores than truly is the case in the distribution of the underlying construct.

Increases in score variability also occurred as in brain-injured groups the range of performances is often greater than in normative groups, particularly when compared across a group of instruments. As not all individuals with traumatic brain injuries have difficulties in every modality, for each domain there may be some that are performing in the normal range, while others are considerably below this. Therefore, this increase in variability may present statistical outliers, particularly in the brain-injured groups with small sample sizes.

Each of the reasons described above may result in scores in the current sample that are statistically unusual, but that are nonetheless valid estimates of that participant's performance[23]. For this reason, the decision was taken to include such data in the analyses. As discussed above, non-parametric statistics were calculated in parallel with parametric statistics in analyses of group differences. These non-parametric statistics are less sensitive to the effects of outliers, so where a genuine outlier was exerting undue influence over a result, the non-parametric statistics could indicate a different finding.

Missing Values

The number of valid cases for each variable, by group and administration format, is presented in Table 18. A number of factors led to measures not being completed by some participants. Firstly, in the TBI group one of the participants drawn from a more remote geographical location completed the first half of the protocol, but a subsequent suitable time to complete the protocol could not be conveniently arranged. The participant in the normative group who completed less than half the protocol refused to complete some measures during her first session and it was decided not to continue with the second half of the assessment.

---

[23] In the case of tests with low ceilings, it may in fact be that the majority of scores are underestimates of the true scores, while the 'outliers' are accurate measurements.

Table 18. Number of valid observations, by group and administration format.

	ED	TBI	Norm
Total N:	10	18	80
Yes/No Response Card			
GSH Orientation Test	10	18	80
VPA Recognition	10	18	79
Logical Memory A and B	10	18	80
Faces A and B	10	18	80
Auditory Reception	10	18	80
Multiple-choice Response Selection			
Verbal Paired Associates	10	18	79
FANL-C Novel and Familiar	10	17	80
Matrix Reasoning	10	18	79
MASC Total Correct and Categories	8	18	80
MASC Perseveration	4	15	76
Number Board			
Graded Attentional Test	10	18	80
Digits Forwards and Backwards	10	18	80
Alphabet Board			
Family Pictures A and B	10	18	80
Boston Naming Test	8	18	80
Hooper VOT	8	18	80
Standard Administration			
Digits Forwards comp.		17	79
Digits Backwards comp.		17	79
AVLT Trials 1-5		17	79
AVLT Delayed Recall and Recognition		17	78
Babcock Story A and B		17	79
Family Pictures A and B comp.		17	78
Boston Naming Test comp.		17	79
Hooper VOT comp.		17	79
Block Design		17	79
WCST Total Correct and Perseveration		17	76
Digit Symbol Coding and Memory		17	79
Symbol Copy		17	79
Finger Tapping		17	75
Valid Cases Listwise[a]	3	13	66

Note:

[a] When the MASC Perseveration score was excluded, valid cases listwise were 6, 15, and 70 for the ED, TBI, and normative groups, respectively.

In both of these cases, however, the data that had been collected were included in the analyses. Wherever possible, missing data were excluded on a pairwise basis, eliminating a case only for the specific analyses for which the data were missing. However, some of the analyses conducted required listwise exclusion of missing data—in these cases, the individuals mentioned above were excluded.

In addition to these two individuals, a small amount of missing data occurred for other participants. After the first two individuals in the ED group had completed the MASC, changes were made to the test administration, and the administration reported in this research is the final version. Therefore, MASC data are not reported for these two individuals. Data are also missing for a number of cases of the MASC Perseveration score, as this index was only calculated if a participant had successfully established the proceeding categories.

Two participants from the ED group could not complete the Boston Naming Test and the Hooper Visual Organization Test, due to the alphabet board requirements of these tests. The issues associated with the difficulty for these participants to make responses for some tests using an alphabet board are discussed later (Chapter 9).

Finally, for four participants in the normative group, Finger Tapping test data were not collected, because the equipment was unavailable. While two full assessment kits were shared between the examiners for this project, only one Finger Tapping apparatus was available, and this was also shared with the clinical neuropsychology service conducted at the Psychology Clinic, Massey University, Palmerston North, and by consultants from this clinic when working in other cities. An additional Halstead-Reitan tapping apparatus was priced at approximately NZ$500.00, an amount considered unjustifiable for the current project's needs. Research has shown that alternative tapping apparatus may produce varying results (Brandon, Chavez, & Bennett, 1986; Coleman, Moberg, Ragland, & Gur, 1997). As data collection began with the Halstead-Reitan apparatus, other options were not pursued.

Test Administration and Delayed Recall Times

Means and standard deviations of test administration times, as well as median values and ranges, are presented in Table A1 (Appendix I, p. 375). Times presented in the table represent the actual administration times of the measures, recorded on a stopwatch, and excluding breaks either between measures or between blocks of tests. Detailed analyses were not conducted with these figures; nonetheless, administration time is an important consideration in the selection of measures for individuals with severe disabilities, and the material may be useful in the future in terms of selecting appropriate measures. A general trend of increased administration time for individuals with expressive disabilities is observed, as expected. Closer examination also reveals that for certain measures this relative increase in administration time was greater than for other measures (for example, see Verbal Paired Associates A, Boston Naming Test, and Familiar and Novel Language Comprehension Test).

Table A2 (p. 378) presents similar statistics for the interval lengths between the first administration of memory measures, and their delayed recall trial. The decision was made to administer the same tests in the interval for all participants, rather than to control interval time. As can be seen in the table, this resulted in statistically significant group differences for four of the seven memory measures. (These differences remained statistically significant for only three of these measures when the non-parametric Kruskal-Wallis statistic was used; see Table A2). In most groups, mean delay intervals were around the targeted 20-minute mark for most tests. However, an examination of the score ranges reveals considerable within-group differences in intervals—the largest being from 5.88 to 46.15 minutes for Logical Memory in the normative group. Interval length clearly may affect performance on memory measures, and raises the question of the extent to which these varying intervals altered performance. However, the decision was made not to complicate the analyses further by adding this factor as a variable. It is difficult to know how this variation compares to other studies. While mean administration times are occasionally reported in research studies, these factors are not routinely examined.

Average testing times for individuals in each group were calculated from these administration times, and are presented in Table 19. It should be noted that while individuals in the ED group had similar administration times to the other groups, they completed only *half* as many measures in the initial assessment. The noticeable decrease in assessment time for the ED group individuals on follow-up assessments was partly due to the particular ED group individuals who were followed up also having shorter average initial assessment times (2 hrs 13 mins) than the whole group. However, there also appears to have been an actual decrease in administration time, which may have been due to familiarity with the testing situation and materials. Not displayed in these figures is the fact that individuals in the ED group had more frequent breaks than other groups, further lengthening total contact time.

Table 19. Test administration time for initial and follow-up assessments, by group.

	Initial Assessment			Follow-up Assessment		
Group	**Mean**	**Std Dev.**	**Max**	**Mean**	**Std Dev.**	**Max**
ED	2 hrs 34 mins	46 mins	3 hrs 54 mins	1 hr 48 mins	20 mins	2 hrs 12 mins
TBI	2 hrs 48 mins	27 mins	3 hrs 36 mins	1 hr 26 mins	15 mins	1 hr 55 mins
Norm	2 hrs 38 mins	24 mins	3 hrs 47 mins	1 hr 17 mins	13 mins	1 hr 56 mins
Total Testing:		287 hrs			93 hrs	
Est. Total Contact:		395 hrs			128 hrs	

Note:
Where administration time was not available for a particular test for an individual, mean group time was substituted for this analysis. Estimated total contact time was based on an additional one hour for initial assessments and half an hour for follow-up assessments. This time included introductory information, consent procedures, and a number of short breaks for participants, and probably underestimates true contact time for the ED and TBI groups.

Table 19 also presents total testing time for initial and follow-up assessments for the whole sample, and estimated total contact time. In all, some 520 hours of face-to-face contact occurred between examiners and participants.

Reliability

Response Reliability

All participants in the TBI and normative groups achieved 100% performance on the response reliability task. In the ED group, two participants correctly selected 23 out of the 28 items, one selected 27 out of 28, and the remainder made no errors at all. All participants clearly performed above chance levels, and were capable of making the kind of responses required for the measures included in the protocol. As this test was designed primarily as a screen for reliability of responding, it was excluded from all further analyses.

Internal Consistency

Internal consistency scores were calculated for all research measures. Reliability scores were calculated across the whole sample, collapsing across participant groups. This was done for two reasons. Firstly, small participant numbers in the two brain-injured groups would have made group analyses questionable. Secondly, and of paramount importance, reliability coefficients are dependent on the range of scores in a sample, with higher reliability estimates being found when a sample contains a wider range of scores on the measure (Franzen, 1989). To partition by groups would be to make the samples being analysed more homogeneous, and thus artificially deflate reliability estimates.

Cronbach's alpha coefficient was calculated where material was of equal difficulty across test items. However, Cronbach's alpha is a lower bound on the reliability of a test, and particularly underestimates reliability in measures where test items are not all of equal difficulty (e.g., tests where items are presented in increasing difficulty, such as Digits Forwards). Therefore, for such tests, odd-even split-half Pearson's *r* reliability coefficients were calculated instead. These internal consistency scores are presented in the first part of Table 20, and are grouped by the types of adaptation made to the measures.

Table 20. Internal consistency and test-retest reliability of research measures.

Measure	Internal Consistency		Test-Retest Reliability	
		N	r	N
Yes/No Response Card				
VPA Recognition	α = .92	107	.84	69
Auditory Recognition	r = .85[a]	108	.89	69
Faces A	α = .78	108	.72	69
Faces B	α = .76	108	.76	69
Logical Memory B	α = .64	108	.63	69
Logical Memory A	α = .54	108	.67	69
GSH Orientation	α = .42	108	.63	69
Multiple-choice Response Selection				
Verbal Paired Associates A	α = .96	107	.90	69
FANL-C Novel	α = .91	107	.96	68
Verbal Paired Associates B	α = .91	107	.85	69
Matrix Reasoning	r = .91	108	.80	69
MASC Total Correct	α = .91	106	.49	67
FANL-C Familiar	α = .89	107	.84	68
MASC Categories	—[c]		.43	67
MASC Perseveration	—[c]		-.03	59[b]
Number Board				
Digits Forwards	r = .88	108	.77	69
Digits Backwards	r = .86	108	.70	69
Graded Attentional Test	r = .79	105	.70	69
Alphabet Board				
Family Pictures A	α = .92	108	.86	69
Boston (all cues)	r = .92	106	.80	69
Boston Naming Test	r = .91	106	.94	69
Family Pictures B	α = .91	108	.86	69
Hooper VOT	r = .79	106	.91	69

Note:

[a] Odd/even split-half correlation used where test items are ordered by increasing difficulty, making alpha reliability calculations inappropriate. Applies to all *r* values.

[b] N=59 for the perseveration score because calculation of this score requires that an individual has established the previous category (i.e., at least 10 correct responses in a row). Some individuals in the sample did not establish one or both of the categories.

[c] Internal consistency could not be calculated for these scales as they are summary scores, and do not consist of specific individual items.

For tests that used the Yes/No response card, reliability coefficients ranged from .42 for the Good Samaritan Hospital Orientation Test to .92 for Verbal Paired Associates Recognition. Measures that required a multiple-choice response selection had coefficients ranging from .89 to .96. The three tests which used the Number Board ranged in internal consistency from .79 to .88, while measures which used the alphabet board had coefficients ranging from .79 for the Hooper Visual Organization Test to .92 for Family Pictures A.

As reliability scores can be directly affected by the number of items in a scale, it was expected that where adapted measures had been shortened, these tests would have lower reliability coefficients than the measures they were based on. To examine whether there were differences in the reliability of these adapted measures from their standard administrations, internal consistency data for research and standard measures from the protocol were compared (see Table 21). Some research measures had been shortened by dividing them in half; using the other half as their comparison measure provided a length- as well as content-comparable measure. Therefore, in terms of evaluating the effects of response card and alphabet board adaptations, these comparison measures provided a more appropriate benchmark, as previously reported reliability studies have been conducted with the full scales.

For the four tests where research and comparison measures were split halves of the original scale, initial comparisons of internal consistency (reported in Table 21) suggested the surprising finding that the adapted measures had substantially higher internal consistency scores. However, as previously discussed, such reliability scores are dependent on the heterogeneity of the sample. The initial analysis of the research measures included participants from the ED group, who, it was hypothesised, would score at a significantly lower level on many of these tests than the other participants (see *Discriminant Validity,* below, for an investigation of this hypothesis). Therefore, the more heterogeneous sample including this group would be expected to produce higher reliability estimates. The comparison measures were not completed by this group. Therefore, that sample may have had reduced variability, lowering reliability estimates. To examine whether this was the explanation for the observed effect, internal consistency coefficients were re-calculated for the research measures using a reduced sample, consisting of only the TBI and normative groups.

Table 21. Internal consistency of research versus comparison measures.

Measure	**All Groups**		**TBI and Norm**	
		N		N
Logical Memory A	α = .54	108	α = .44	98
Logical Memory B	α = .64	108	α = .63	98
Babcock Story A			α = .72	96
Babcock Story B			α = .87	96
Digits Forwards	r = .88[a]	108	r = .82	98
Digits Backwards	r = .86	108	r = .81	98
Digits Forwards comp.			r = .83	96
Digits Backwards comp.			r = .83	96
Boston Naming Test	r = .91	106	r = .85	98
Boston (all cues)	r = .92	106	r = .86	98
Boston Naming Test comp.			r = .83	96
Boston (all cues) comp.			r = .83	96
Family Pictures A	α = .92	108	α = .87	98
Family Pictures B	α = .91	108	α = .88	98
Family Pictures A comp.			α = .79	95
Family Pictures B comp.			α = .78	95
Hooper VOT	r = .79	106	r = .49	98
Hooper VOT comp.			r = .51	96

Note:

[a] Odd/even split-half correlations used where tests were ordered by increasing item difficulty, making alpha coefficient calculations inappropriate. Applies to all *r* values.

When these new results were examined, a different picture emerged (see Table 21). Reliability coefficients for Digits Forwards and Backwards, the Boston Naming Test, and the Hooper Visual Organization Test fell to levels similar to their comparison measures. This change was most marked for the Hooper Visual Organization Test, which dropped from r=.79 to r=.49. For Family Pictures a smaller change did reduce the gap, but considerable differences remained. For Logical Memory A and B (the tests least like their comparison measures, both in content and reliability coefficients), the gap was widened, with Logical Memory A falling from an already low alpha coefficient of .54 down to .44. Logical Memory A and B remained the only tests with reliability coefficients considerably below their comparison measures.

Test-Retest Reliability

Cicchetti (1994) argued for the use of intraclass correlation coefficients in the examination of both inter-observer and test-retest reliability calculations. Intraclass coefficients evaluate not merely the covariance of scores, but their absolute agreement. While this is critical in inter-observer reliability, the question of whether covariance or absolute agreement is the more suitable criterion in test-retest reliability is debatable. Studies have shown that people perform at significantly different levels on various neuropsychological measures at re-test (e.g., McCaffrey, Ortega, & Haase, 1993; McCaffrey, Ortega, Orsillo, Nelles, & Haase, 1992; McCaffrey & Westervelt, 1995; Slick et al., 1996). This may be due to a number of factors, including recall of the test material, test requirements, or both, or genuine changes in the underlying construct. In this research, test-retest reliability was assessed to determine which tests were stable measures of a specific construct, rather than the temporal stability of test scores. Therefore, the appropriate statistic was Pearson's product-moment correlation coefficient.

Pearson's r test-retest reliability coefficients were calculated for research measures, and are presented in the second part of Table 20 (p. 169). Coefficients ranged from .63 to .89 for measures using the Yes/No response card. For measures using multiple-choice response selections, test-retest reliability ranged from .80 for Matrix Reasoning to .96 for the Familiar and Novel Language Comprehension Novel items score, with the exception of the Match and Shift Categories test, which had reliability coefficients of .43 and .49 for the Categories and Total Correct scores, and the only statistically non-significant test-retest reliability of -.03 for the Perseveration score.

Tests using the number board had reliability coefficients of .70 for the Graded Attentional Test and Digits Backwards, and .77 for Digits Forwards. The highest average test-retest reliability coefficients across a category were for measures using the Alphabet Board, ranging from .80 for the Boston Naming Test (all cues) score, to .94 for the Boston Naming Test.

Validity

Concurrent Validity

To examine the extent to which adapted measures examined a similar construct to their comparison measures, Pearson's *r* correlation coefficients were calculated between the research and comparison measures and are presented in Table 22. As the comparison measures were administered only to the TBI and normative groups (since ED group individuals were physically incapable of completing these tests), the analysis is based on these two groups. In addition to point estimates for *r*, 95% confidence intervals are reported, which were calculated using Fisher's *r* to *z* conversion, and computed using the engine at the Vassar College web site (http://faculty.vassar.edu/%7Elowry/rho.html). With the exception of the correlations between the Match and Shift Categories test and the Wisconsin Card Sorting Test, all relationships were statistically significant ($p < .001$). Of the significant relationships found, correlations ranged from .45 for the Hooper Visual Organization Test and comparison measure to .91 for each of the scores calculated from the Boston Naming Test. Scatterplots for each pair of variables in the table did not indicate any non-linear relationships.

The earlier results had indicated that the tests in the research battery had varying degrees of (un)reliability (see Table 20 and Table 21). Therefore, correlations corrected for attenuation due to this unreliability are also reported in Table 22. This correction was suggested by Schmitt (1996)[24]. Theoretically, these corrected reliability coefficients represent the relationship that would be observed between the two variables, were they both perfectly reliable measures.

---

[24] $$r_{ABcorr} = \frac{r_{AB}}{\sqrt{rel_A} \times \sqrt{rel_B}}$$

Where

r_{ABcorr} = Correlation coefficient between A and B corrected for unreliability.

r_{AB} = Correlation coefficient between A and B.

rel_A = Reliability coefficient for variable A.

rel_B = Reliability coefficient for variable B.

Table 22. Correlations between research and comparison measures.

Research — Comparison Measures	r	95% C.I.	r^2	r_{corr}	95% C.I.	r_{corr}^2	N
Yes/No Response Card							
VPA Recognition — AVLT[a] Recognition	.61**	(.47 to .72)	37%				95
Logical Memory B — Babcock Story B	.58**	(.43 to .70)	34%	.78	(.69 to .85)	61%	96
Logical Memory A — Babcock Story A	.47**	(.30 to .61)	22%	.84	(.77 to .89)	70%	96
Multiple-choice Response Selection							
Verbal Paired Ass. B — AVLT Delayed Recall	.75**	(.65 to .83)	56%				95
Verbal Paired Ass. A — AVLT Trials 1-5	.67**	(.53 to .77)	45%				96
MASC Categories — WCST Total Correct	.02	(-.18 to .22)	0%				93
MASC Total Correct — WCST Total Correct	.01	(-.19 to .21)	0%				93
MASC Perseveration — WCST Perseveration	-.16	(-.36 to .05)	3%				86
Number Board							
Digits Backwards — comparison	.80**	(.72 to .86)	64%	.98	(.97 to .99)	95%	96
Digits Forwards — comparison	.78**	(.69 to .85)	61%	.95	(.93 to .97)	89%	96
Alphabet Board							
Boston Naming Test — comparison	.91**	(.87 to .94)	83%	1.00[b]		100%	96
Boston (all cues) — comparison	.91**	(.87 to .94)	83%	1.00[b]		100%	96
Family Pictures B — comparison	.62**	(.45 to .73)	38%	.75	(.65 to .83)	56%	95
Family Pictures A — comparison	.58**	(.43 to .70)	34%	.70	(.58 to .79)	49%	95
Hooper VOT — comparison	.45**	(.28 to .60)	20%	.90	(.85 to .93)	81%	96

Note:

[a] Only total scores for each trial were coded for the AVLT. As a result, internal consistency data could not be coded, and therefore r_{corr} values are not reported.

[b] Calculated r_{corr} value = 1.08. As true correlations cannot exceed 1.00, these are rounded down. Issues relating to the potential for error in correcting correlations for unreliability are discussed in the following chapter. Confidence intervals for correlations of 1.00 do not deviate from 1.00 and so are not reported.

** $p < .001$. All other correlations $p > .10$.

Correlations corrected for unreliability (r_{corr}) are reported where internal consistency coefficients could been calculated. Correlations reported here are between the TBI and normative groups, and so reliability coefficients calculated from these groups were used (see Table 21).

However, there are problems associated with the practice of correcting correlations for unreliability. Therefore, in line with Schmitt's (1996) guidelines, both corrected and uncorrected correlations are reported, along with r^2 values. These latter values provide an estimate of the percentage of shared variance, for the two measures of relationship. The more these values deviate from 100%, the greater the likelihood that test modifications have substantially altered the construct being tapped by the measure[25]. For the variables that had statistically significant relationships, uncorrected r^2 values ranged from 22% to 83%, and corrected r^2 values ranged from 49% to 100%.

Construct Validity

In addition to correlations between research measures and their direct comparisons, it was expected that statistically significant relationships would be observed between a range of measures. The large number of bivariate correlations which could have been calculated from the variables in the study once again raised the issue of Type I error inflation, and, in addition, would pose significant difficulties in terms of understanding and reporting the observed relationships.

One parsimonious method for examining complex interrelationships between variables is factor analysis, and an exploratory factor analysis was undertaken to investigate whether variables loaded on meaningful factors. If research and their comparison measures loaded together in clearly defined factors this would provide further evidence for the construct validity of the adapted measures.

As the purpose of this analysis was to identify shared variance between variables, Common Factor Analysis (known as Principal Axis Factoring in SPSS) was the appropriate extraction method (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Domains of neuropsychological functioning have been shown to interrelate in the past, and it was expected that factors would be correlated; therefore, the oblique OBLIMIN factor rotation technique was employed, as it is the appropriate technique to use under such circumstances.

---

[25] There are a number of other possible explanations for deviations from a perfect relationship between the research and comparison measures. See Chapter 9, *Discussion*, for an examination of these issues and those associated with correcting correlations for unreliability.

This factor analysis had a cases:variables ratio lower than 5.0, and considerably lower than the 10 cases per variable recommended by some authors in the past. While omitting some variables was considered, this omission might have reduced the overdetermination of the factors, yet not necessarily reduce the number of actual factors present in the data. This course of action was therefore considered potentially counterproductive, since overdetermination improves the likelihood of arriving at a stable and accurate factor structure. Furthermore, recent findings indicate that the cases-to-variables ratio is not the only issue which must be considered, with the degree of communality in the sample interacting with both sample size and this ratio (MacCallum, Widaman, Zhang, & Hong, 1999). Therefore, within the limitations of the sample size the analysis was considered exploratory, but of potential use if a meaningful factor structure could be uncovered.

Initial factor extraction indicated two strong factors, which accounted for 39% of the total variance in the sample. Scree plot analysis suggested that between three and seven factors in total should be extracted. Previous research has shown that overextraction results in more accurate factor structure solutions, while underextraction can produce significant errors (Wood, Tataryn, & Gorsuch, 1999). While not fitting exactly with the number of domains that the protocol measures had previously been divided into (see Table 10 and Table 11, Chapter 6), seven factors was generally consistent with the hypothesised constructs. Therefore, the analysis was re-run, and seven factors were initially extracted.

When this initial solution was rotated, Finger Tapping, the MASC Perseveration score and the Familiar and Novel Language Comprehension Familiar Language score did not have loadings above .30 on any of the factors. Furthermore, as the first two of these variables had a high proportion of missing data, their inclusion caused 10 cases to be excluded from the analysis. Therefore, these first two variables were removed, and the analysis re-run.

The results of this reduced data set suggested that six, rather than seven, factors should be extracted. This analysis was conducted, the factors rotated, and the resulting factor loadings are presented in Table 23.

Table 23. Factor structure of research and comparison measures.

Variables	**Factor Loadings**						**Communalities**
	1	**2**	**3**	**4**	**5**	**6**	
Babcock Story B	**0.83**						0.77
AVLT Trials 1-5	**0.83**						0.79
AVLT Delayed Recall	**0.78**	-0.36					0.81
Digit Symbol Memory	**0.78**						0.64
Verbal Paired Associates B	**0.74**	-0.33					0.72
Verbal Paired Associates A	**0.74**						0.65
Family Pictures B	**0.72**	-0.35					0.81
Block Design	**0.71**						0.59
AVLT Recognition	**0.69**	-0.36					0.65
Family Pictures A	**0.67**	-0.34					0.73
Family Pictures B comp.	**0.65**	-0.33					0.60
Logical Memory B	**0.63**						0.42
Family Pictures A comp.	**0.62**	-0.31					0.59
Hooper VOT comp.	**0.59**						0.44
VPA Recognition	**0.58**						0.49
Faces A	**0.57**						0.48
Faces B	**0.56**						0.39
Logical Memory A	**0.55**						0.34
Matrix Reasoning	**0.53**						0.32
WCST Perseveration	**-0.51**						0.35
GSH Orientation	0.42						0.36
Graded Attentional Test	0.42						0.22
Boston (all cues) comp.	**0.49**	**0.77**					0.91
Boston (all cues)	**0.54**	**0.70**					0.84
Boston Naming Test	**0.55**	**0.69**					0.81
Boston Naming Test comp.	**0.54**	**0.66**					0.76
Auditory Reception	0.35	**0.69**					0.63
FANL-C Novel	0.46	**0.60**					0.61
FANL-C Familiar	0.48	0.39	-0.30				0.56
Digits Backwards	**0.58**		**0.59**	-0.38			0.83
Digits Backwards comp.	**0.53**		**0.55**				0.69
Digits Forwards comp.	**0.61**		**0.52**				0.75
Digits Forwards	**0.59**		**0.51**				0.72
MASC Total Correct	0.49		-0.38	**-0.63**			0.82
MASC Categories	**-0.54**		0.40	**0.61**			0.86
Symbol Copy	**0.51**				-0.42		0.51
Digit Symbol Coding	**0.63**				-0.37		0.66
WCST Total Correct		-0.32			0.32		0.38
Babcock Story A	**0.64**					0.42	0.67
Hooper VOT	**0.58**					-0.33	0.51
Shared Variance Explained:	36.0%	11.1%	5.2%	4.1%	2.7%	2.5%	

Note:
Correlations over ± .30 are presented in the table, and those over ± .50 are in bold.

In this reduced data set, the Familiar and Novel Language Comprehension Familiar Language score now had loadings of greater than .30 on three factors. Variable communalities, reported in the table, ranged from just .21 for the Graded Attentional Test, to .90 for the Boston Naming Test comparison measure. This range of communalities is 'wide' by criteria used in one previous study (MacCallum et al., 1999), which found that data sets with wide-ranging communalities produced factor solutions significantly more congruent than where only low communality was observed. Such samples were also less susceptible to the effects of smaller sample sizes and low cases:variables ratios than samples with low communalities[26].

Table 23 also shows the proportion of shared variance accounted for by each of the six factors. These factors together accounted for almost 61.6% of the shared variance in the sample. Of this, over half was accounted for by the first factor. With the exception of the Wisconsin Card Sorting Test Total Correct score, every variable had a loading over .30 on this first factor. Loading most strongly were the memory measures (which by sheer numbers dominated the protocol). However, due to the almost universal loading on this factor it was labelled 'General Cognitive Functioning'. Due to the number of variables that loaded on this factor, variables that had a strong and logical loading on another factor were listed with that factor in the table, even if the higher coefficient was with the first factor.

The second factor was composed of the seven language measures, and was labelled 'Language'. Three verbal memory measures and all four of the Family Pictures measures, and the Wisconsin Card Sorting Test Total Correct score all had significant negative loadings on this factor.

The third factor was similarly strongly characterised, and was labelled 'Memory Span', the highest loading variables being the four Digits Forwards and Digits Backwards measures.

---

[26] In cases of small samples and low variables:cases ratios, however, samples with exclusively high communalities produce more congruent results again (MacCallum et al., 1999).

The remaining three factors were less clearly identifiable. The only loadings on Factor 4 were the two remaining Match and Shift Categories Test variables included in the analysis, with a negative loading from Digits Backwards. Higher scores on MASC Categories indicate lower performance, resulting in the negative loadings of the other variables. It is possible that this factor represents the Working Memory aspects of both Digits Backwards and the Match and Shift Categories test.However, the interrelationships between these variables cannot be described simply. An examination of Factors 3 and 4 reveals that the MASC variables load in the same direction on both factors, whereas the Digits Backwards test reverses its loading between factors.

Neither was Factor 5 entirely clear, with Symbol Copy and Digit Symbol Coding having negative loadings on this factor, while the Wisconsin Card Sorting Test Total Correct score had a positive loading. This indicates that there was a trend to make a greater number of correct answers on the Wisconsin if one was slower on the measures of information processing and motor speed. It is plausible that this factor could therefore represent an 'Impulsiveness/Caution' factor, or could indicate a speed-accuracy trade-off; however, these are offered only as tentative hypotheses.

The final factor had loadings over .30 from only two variables, the Babcock Story Recall A (.42), and the Hooper Visual Organization Test (-.32). No clear relationship between these variables could be identified, and no label is offered for this factor.

The results of the factor analysis provide some evidence for the construct validity of the measures. An overall cognitive functioning factor was identified, which corresponds with previous theory and research regarding *g* and the concept of IQ. At their most basic, the adapted measures are intended to tap cognitive functions, and therefore should load on this factor, as was observed.

The language factor was the most clearly characterised, and provides strong support for the validity of the adapted language measures. Similarly, loadings on the third factor suggest that the adapted Digit Span measures tap a similar or related construct to their standard counterparts. Remaining factors are not clear, and do not appear to provide further evidence that speaks either way to the issue of the construct validity of the research measures.

Discriminant Validity

Effects of Demographic Variables on Test Performance

Despite efforts to recruit samples equivalent in terms of demographic factors, the two brain injured groups differed significantly from the normative group in terms of gender, $\chi^2(2)=12.47$, $p=.002$, and education, $\chi^2(3)=12.89$, $p=.005$, but not age group, $\chi^2(3)=2.00$, n.s.[27]

These differences do not necessarily reflect inadequacies in sampling techniques. Males tend to be over-represented in traumatically brain-injured populations, and level of education is often linked to socio-economic status, which is also related to traumatic brain injury (Naugle, 1990), suggesting the current sample may be representative of the traumatically brain injured population. However, it is arguable whether in most cases a normative sample should include numbers of males and females in proportion to their representation in the general population[28].

As demographic variables were confounded with group, they could alter the results of an examination of group differences if they were also related to test performance. Cicchetti (1994) described systematic stratification of intelligence tests on a range of demographic variables as a "critical process [that] cannot be underestimated" (p. 284). However, some previous research has found that while demographic variables such as age and education may affect test performance in normative

---

[27] For these analyses, the ED and TBI groups were collapsed together to provide sufficient expected values in all cells. In order to include age and level of education in the χ^2 these were transformed into categorical variables. Participants were separated into four age group categories, divided by the median, upper and lower quartile values calculated from the normative group. Education was similarly divided into four groups, consisting of: three or fewer years of secondary education; four to five years of secondary education; the completion of a tertiary diploma or part of a degree; and the completion of a degree or postgraduate qualification.

[28] In fact, normative samples should be selected to closely match the population with which the measure will be used. However, it is usually not possible to predict *a priori* the groups with which a test may subsequently be employed. As all such samples are subsets of the general population, a group representative of the general population is the best 'average' sample. Of course, wherever possible these should be of sufficient size that normative data can be stratified by the relevant variables (e.g., gender, age, and education level), allowing future direct comparison to the subset of interest.

groups, they may not alter performance of those with traumatic brain injuries in the same way, if at all (Reitan & Wolfson, 1995). If this were the case, any traditional statistical control process (e.g., covariance) would be inappropriate to use in this study, as it would not take account of this interaction between brain injury status and the effects of demographic variables. Furthermore, since these demographic variables were confounded with group, an examination of the effect of the demographic variables would be affected by group status. Due to small cell sizes across these demographic variables within groups, there is not sufficient statistical power to control for their effects, even if it was desirable to do so.

It was decided that there was no statistically appropriate remedy to this situation. Therefore, having noted that there were demographic differences between the groups, no additional steps were taken to control for these effects in the analyses. The one exception to this was that analyses were re-run including only the males in the sample, to determine whether this altered the results[29]. That other controls were not possible is acknowledged as being far from ideal, but as the study had not been designed to investigate these issues, limitations in the current sample prevented a more comprehensive examination. The following results should consequently be interpreted cautiously. It can also be noted that such effects could only affect the current examination of group differences; in all other analyses, the investigation was of within-individual relationships in test scores, and therefore participants acted as their own controls.

Group Differences In Test Performance

To examine whether there were any significant effects of group on research measure performance, an omnibus MANOVA was first calculated with group as the independent variable. A significant effect of group was found, Wilks' λ=0.056, F(46, 134)=9.43, $p < .001$, ES=0.76[30]. However, due to listwise deletion of cases in MANOVA analysis, this calculation was based on only part of the dataset. That is, MANOVA requires that all cases have values for all variables. Any case missing

---

[29] There were insufficient women in the brain-injured groups to run a separate analysis of the female participants also.

[30] Effect size (ES) for MANOVAs calculated by SPSS 6.1.1 for Macintosh.

data for even a single variable is excluded. In the current analyses this requirement excluded seven of the ten individuals in the ED group, and four of the eighteen people in the TBI group[31]. Therefore, in order to maximise the number of cases included, individual ANOVAs were then conducted to examine the effects of group on individual measures. Where significant effects were found for a variable, two pairwise comparisons were planned, in line with earlier hypotheses: the first, of the ED group with the TBI group, and the second, of the TBI group with the normative group.

Table 24 displays the results of the ANOVAs examining the sensitivity of research measures to group differences. With the exception of the MASC Perseveration score, for which there was no significant difference between groups, significant differences were found between groups on all measures. An earlier examination of the distributions of these variables had indicated some may have violated the assumptions of ANOVA, particularly that of an underlying normal distribution. Therefore, the non-parametric Kruskal-Wallis One-Way ANOVA was used in parallel with the parametric analyses. No differences were found in the Kruskal-Wallis results compared to the parametric ANOVA; therefore, only parametric statistics are reported.

Analyses were also conducted on the sub-sample of males only. One result differed in this re-analysis, with the MASC Total Correct score remaining significant with the parametric statistic, $F(2, 55)= 6.125, p=.004, f= 0.47$, but becoming marginally non-significant with the Kruskal-Wallis test, $\chi^2(2)=4.91$, $p=.086$. In light of the reduced power of this statistic over the parametric result, the trend towards significance ($p < .10$) and the positive findings on both tests with the full sample, evidence still points to real differences between the groups.

---

[31] The loss of this many cases in the two clinical groups is clear cause for concern, and therefore the results of analyses conducted on this reduced data set should be interpreted with a great deal of caution. Nonetheless, a finding of statistically significant differences in such an analysis, even with this loss of statistical power provides an indication of genuine group differences.

Cohen's *f* effect size statistic was calculated for all variables, using the formulae from the American Psychological Association (1994) and Cohen (1977, 1988) for effect sizes in Analysis of Variance.[32] The statistic *d* is calculated by $d = 2f$, and is presented in the table to allow direct comparison with effect sizes presented in further analyses below. Other than MASC Perseveration, all variables meet Cohen's (1977, 1988) criteria for large effects, with effect size greater than 0.50.

Table 24. Group differences in performance on research measures.

Variable	**F-statistic**	***f***	***d***
GSH Orientation	F(2, 105)=13.28	0.50	1.00
Graded Attentional Test	F(2, 105)=47.83	0.95	1.90
Digits Forwards	F(2, 105)=24.84	0.69	1.38
Digits Backwards	F(2, 105)=33.10	0.79	1.58
Verbal Paired Associates A	F(2, 104)=72.61	1.18	2.36
Verbal Paired Associates B	F(2, 104)=102.11	1.40	1.80
VPA Recognition	F(2, 104)=89.27	1.31	2.62
Logical Memory A	F(2, 105)=27.67	0.73	1.46
Logical Memory B	F(2, 105)=20.76	0.63	1.26
Faces A	F(2, 105)=41.69	0.89	1.78
Faces B	F(2, 105)=39.53	0.87	1.74
Family Pictures A	F(2, 105)=91.33	1.32	2.64
Family Pictures B	F(2, 105)=73.14	1.18	1.36
Auditory Reception	F(2, 105)=29.27	0.75	1.50
Boston Naming Test	F(2, 103)=30.70	0.77	1.54
Boston (all cues)	F(2, 103)=26.77	0.72	1.44
FANL-C Novel	F(2, 104)=85.88	1.29	2.58
FANL-C Familiar	F(2, 104)=107.47	1.44	2.88
Hooper VOT	F(2, 103)=93.50	1.35	1.70
Matrix Reasoning	F(2, 104)=49.48	0.98	1.96
MASC Total Correct	F(2, 103)=15.66	0.55	1.10
MASC Categories	F(2, 103)=12.80	0.50	1.00
MASC Perseveration	F(2, 92)=0.73[a]	0.13	0.26

Note:

[a] p=.487. All other F-values are $p < .001$.

---

[32] $$\eta^2 = \frac{v_1 \cdot F}{(v_1 \cdot F) + v_2}, \text{ and } f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

Where

η^2 = Eta squared. F = F-statistic.

v_1 = Numerator degrees of freedom. v_2 = Denominator degrees of freedom.

f = f effect size statistic.

One-tailed t-tests were conducted to examine where significant differences lay in the samples for each of the variables reported as having significant differences (see Table 24). The results of these t-tests are presented in Table 25. As some variables were known to have markedly non-normal distributions, Levene's test for equality of variances was conducted. Where variances were significantly different (α=.05), the unequal variances t-test was used to compare groups. In only two cases did this procedure change the outcome of a test, with differences between the TBI and normative groups on the Good Samaritan Hospital Orientation Test and the Graded Attentional Test becoming non-significant using the unequal variances t-test. For the majority of tests, significant differences were found between both the ED and TBI groups, and the TBI and normative groups.

Additionally, the non-parametric Mann-Whitney U statistic was calculated to examine whether non-normal distributions may have altered the results of this analysis. For all but one variable this test supported the findings of the t-tests, and where the results did not differ, the parametric results are reported. However, for the Good Samaritan Hospital Orientation Test, both ED–TBI group differences, and TBI–Norm group differences became significant using this test (see Table 25). For the TBI–Norm group comparison, this finding thus reversed the change made when the unequal variances t-test was used.

The effect size, *d,* statistic was calculated for each t-test[33]. The value of *d* can be interpreted as the number of standard deviations between the group means, and thus provides a readily interpretable index of the strength of the difference between the groups. For comparisons between the ED group and TBI group, values of *d* ranged

---

[33] $d = \frac{(Y_a - Y_b)}{\sqrt{V_w}}$, and $V_W = \frac{(n_a - 1)V_a + (n_b - 1)V_b}{(n_a - 1) + (n_b - 1)}$

Where

Y_a = Mean of group a.	n_a = Sample size for group a.
Y_b = Mean of group b.	n_b = Sample size for group b.
V_a = Variance of group a.	V_w = Pooled variance for the sample.
V_b = Variance of group b.	

from 0.44 to 3.34, clearly indicating the ED group performed at clinically significant levels below the TBI group on a range of measures. Similarly, comparisons of the TBI group and the normative group revealed *d* values up to 2.15, with all effect sizes except the statistically non-significant Auditory Reception and FANL-C Novel scores being greater than 0.45. In both group comparisons, the difference between group means was over one standard deviation across the majority of variables. The analysis provides strong evidence that these measures are sensitive to the effects of traumatic brain injury, and that individuals in the ED group performed, both statistically and clinically, at significantly lower levels on these measures.

Table 25. Paired comparisons of effects of group on research measure performance.

Variable	**ED vs. TBI**		**d**	**TBI vs. Norm**		**d**
GSH Orientation	t(26)=1.63 U(18,10)=61	n.s. p=.040	0.64	t(17.83)=1.43[a] U(80,18)=632	n.s. p=.035	0.66
Graded Attentional Test	t(26)=4.93	p < .001	1.95	t(20.21)=1.58[a]	n.s.	0.54
Digits Forwards	t(26)=3.18	p =.002	1.25	t(96)=3.83	p < .001	1.00
Digits Backwards	t(26)=4.75	p < .001	1.87	t(46.94)=6.45[a]	p < .001	1.17
Verbal Paired Ass. A	t(22.32)=5.34[a]	p < .001	1.65	t(95)=6.86	p < .001	1.79
Verbal Paired Ass. B	t(22.74)=4.43[a]	p < .001	1.37	t(18.60)=5.25[a]	p < .001	2.15
VPA Recognition	t(11.30)=3.94[a]	p =.001	1.88	t(17.00)=2.25[a]	p =.019	1.26
Logical Memory A	t(26)=2.68	p =.007	1.06	t(20.98)=3.11[a]	p =.003	1.01
Logical Memory B	t(26)=1.11	n.s.	0.44	t(20.00)=3.32[a]	p =.002	1.16
Faces A	t(26)=2.42	p =.012	0.96	t(96)=6.22	p < .001	1.62
Faces B	t(26)=4.25	p < .001	1.68	t(96)=4.67	p < .001	1.22
Family Pictures A	t(25.06)=6.70[a]	p < .001	2.16	t(96)=7.63	p < .001	1.99
Family Pictures B	t(25.06)=4.43[a]	p < .001	1.43	t(96)=7.54	p < .001	1.97
Auditory Reception	t(10.07)=3.53[a]	p =.003	1.78	t(96)=0.12	n.s.	0.03
Boston Naming Test	t(24)=3.92	p < .001	1.66	t(96)=2.55	p =.006	0.67
Boston (all cues)	t(8.81)=2.53[a]	p =.017	1.35	t(96)=2.70	p =.004	0.71
FANL-C Novel	t(9.90)=6.64[a]	p < .001	3.34	t(95)=0.54	n.s.	0.14
FANL-C Familiar	t(9.72)=5.15[a]	p < .001	2.61	t(95)=1.71	p =.046	0.46
Hooper VOT	t(24)=7.30	p < .001	3.10	t(96)=4.15	p < .001	1.08
Matrix Reasoning	t(26)=6.16	p < .001	2.43	t(95)=3.33	p < .001	0.87
MASC Total Correct	t(24)=2.73	p =.006	1.16	t(96)=1.90	p =.030	0.49
MASC Categories	t(24)=-2.06[b]	p =.025	-0.88	t(20.43)=-1.78[a, b]	p =.045	-0.60

Note:

[a] Levene's test for equality of variances indicated significant differences of variances for this test, so the unequal variances t-test is reported.

[b] t-value negative as higher scores were related to lower performance on this scale.

The effect sizes for the group comparisons on the Good Samaritan Hospital Orientation test (*d*=0.64 and 0.66) were large by Cohen's (1977, 1988) criterion. This supported the finding of differences between the groups from the non-parametric statistic.

Normative Data

Based on the performance of the normative group, it was planned to produce normative data against which clinicians could compare the performance of clients, should they wish to use the adaptations reported in this research. At its most basic, normative data may consist simply of mean scores and standard deviations for a normative population, a range of clinical groups, or preferably both. However, such scores assume an underlying normal distribution, and may markedly skew interpretations of test performance if this assumption is violated. An example of this, particularly relevant to the current research, would be a measure which attempts to tap an ability which is normally distributed in the general population, but using a scale which has a low ceiling. Thus, the majority of individuals in the normative group will score at the top of the scale. In such a distribution, norms must be reported in another form. Percentile data are an alternative that provide an accurate picture of the distribution in the healthy population. However, such a scale is problematic because it does not have equal intervals, complicating interpretation. That is, the clinical significance of a 10-point difference at the extremes of the scale is greater than of a 10-point difference around the mean.

Another approach is to standardise raw scores on the measure, converting them to a scale with known properties and a normal distribution. The most widely used example of this approach is the Wechsler Intelligence Scales, which convert raw scores to a scale from 1 to 19, with a mean of 10 and a standard deviation of 3. Such conversion allows not just easy comprehension of individual test scores, but meaningful comparisons between tests with different raw data scales and scoring methods.

Data collected from the normative group in this study exhibited both substantial departures from normality, and significant ceiling effects. Anticipating this, the decision was made to convert raw scores to a standardised scoring system similar to that used in the Wechsler Scales, with a mean of 10, and a standard deviation of 3. Such standardised scores would be immediately assessable to practitioners, being presented in a familiar format, and would have the benefit, discussed above, of allowing easy comparison between tests.

Based on the normative group data, standard score equivalents for raw scores on the research measures were developed. The appropriate standard score to be allocated to each raw score was calculated using a procedure designed to produce the closest possible fit for the desired distribution (i.e., mean of 10 and standard deviation of 3) within the limitations of the observed data. The procedure involved matching percentile ranges in the observed distribution to those in a normal distribution with the required mean and standard deviation. For raw data values which fell within the observed distribution *range*, but which were not observed in the current normative group, standard score equivalents were interpolated. The procedure by which standard scores were calculated and interpolated was not mathematically complex, but the description of it is sufficiently convoluted that it is presented in Appendix J (p. 379).

Based on the calculated score equivalents, tables for converting raw scores on the research measures into standardised scores were developed, and are presented in Table 26[34]. Raw test scores obtained by participants on the research measures were then converted into standardised scores. The distribution of these scores in the normative group was examined to investigate the extent to which the desired normal distribution had been produced. The means and standard deviations of the standardised variables in the normative group are presented in Table 27. A Kolmogorov-Smirnov test of normality was also conducted for each of these variables, and the results are also presented in Table 27.

---

[34] Standardised score equivalents for the Good Samaritan Hospital Orientation Test and Verbal Paired Associates Recognition were not calculated since the normative group had perfect performance on the second, and near-perfect on the first.

Table 26. Standard score conversion tables for research measures.

	GAT	DF	DB	VPA	VPB	LMA	LMB	FA	FB	FPA	FPB
19											
18											
17								46-48			
16		16	14-16					45	45-48		32
15			13					44	44	32	
14		14-15	12	32		15	15	43	43		31
13			11					41-42	42	30-31	30
12	7	13	10	31			14	39-40	40-41	29	29
11		12	9	30	8	14		37-38	39	28	28
10		11	8	29			13	35-36	38	27	27
9	6	10	7	26-28		13	12	34	36-37	25-26	25-26
8		9		23-25	7		11	32-33	34-35	22-24	21-24
7	5	8	6	20-22	6	12		31	32-33	20-21	20
6			5	14-19	5	11	10	29-30	31	17-19	17-19
5	4		4	12-13	4		9	27-28	29-30	15-16	15-16
4		7		9-11		10		26	27-28	14	13-14
3	≤ 3	≤ 6	≤ 3	≤ 8	≤ 3	≤ 9	≤ 8	≤ 25	≤ 26	≤ 13	≤ 12
2											
1											

	AR	BNT	BNT (all cues)	FANL-C Novel	FANL-C Familiar	VOT[a]	MR	MASC Total	MASC Cat.	MASC Pers.
19										
18										
17							26	40	20	≤ 7
16							25			
15									21	8
14		30				30	24	39		9
13						29	23		22	
12	50			20		28	22	38		
11		29	30		20	27	21		23	10
10	49	28				26	20	37	24	
9		27		19		25	18-19	36	25-26	
8	48	26	29			24	17	34-35	27-29	
7	47	25	28	17-18	19	23	15-16	32-33	30-34	11
6	45-46	23-24	27	16		22	13-14	28-31	35-42	
5	44	20-22	26	14 - 15		21	10-12	24-27	43-46	
4	39-43	16-19	2 -25	10 - 13	16-18	19-20	8-9	21-23	47-49	≥ 12
3	≤ 38	≤ 15	≤ 22	≤ 9	≤ 15	≤ 18	≤ 7	≤ 20	≥ 50	
2										
1										

Note:

[a] See Limitations of the Research, Chapter 9, p. 214 for a caution in the use of this measure.

Abbreviations used in the table: GAT - Graded Attentional Test; DF - Digits Forwards; DB - Digits Backwards; VPA & VPB - Verbal Paired Associates A and B; LMA & LMB - Logical Memory A and B; FA & FB - Faces A and B; FPA & FPB - Family Pictures A and B; AR - Auditory Reception; BNT - Boston Naming Test; FANL-C - Familiar and Novel Language Comprehension test; VOT - Hooper Visual Organization Test; MR - Matrix Reasoning; MASC - Match and Shift Categories Test; Cat. - Categories; Pers. - Perseveration. Data are from adapted research versions.

Table 27. Distribution of standard scores in the normative group.

Standardised Variable	Mean	Standard Deviation	Kolmogorov-Smirnov	
Graded Attentional Test	10.07	2.39	3.04	p < .001
Digits Forwards	9.94	2.79	1.15	n.s.
Digits Backwards	10.20	2.88	1.16	n.s.
Verbal Paired Associates A	10.14	2.91	1.20	n.s.
Verbal Paired Associates B	9.97	2.01	4.16	p < .001
Logical Memory A	10.06	2.85	1.38	p =.045
Logical Memory B	9.82	2.75	1.32	p =.061, n.s.
Faces A	10.07	3.01	0.75	n.s.
Faces B	10.02	2.90	1.01	n.s.
Family Pictures A	10.18	3.20	0.83	n.s.
Family Pictures B	10.02	3.08	0.83	n.s.
Auditory Reception	9.76	2.47	2.13	p < .001
Boston Naming Test	9.95	2.71	1.45	p =.030
Boston (all cues)	9.52	2.12	3.52	p < .001
FANL-C Novel	9.84	2.43	2.69	p < .001
FANL-C Familiar	9.99	2.07	4.25	p < .001
Hooper VOT	9.84	2.81	1.38	p =.044
Matrix Reasoning	10.21	2.90	0.94	n.s.
MASC Total Correct	9.85	2.75	1.19	n.s.
MASC Categories	9.85	3.06	1.04	n.s.
MASC Perseveration	10.12	2.89	2.42	p < .001

For the majority of test scores, the mean fell within ±0.20 of the desired score of 10, with the means for Auditory Reception (9.76) and the Boston Naming Test (all cues; 9.52) falling below this range, and Matrix Reasoning (10.21) above. Standard deviations tended to be further from the target, with few above 3.0, but many falling well short of this. This reduced variance in the sample is attributable to the tendency for scores of participants in the normative group to cluster at the ceilings of many of the measures. Additionally, as many of the research measures were shortened, the range of possible scores was also reduced, resulting in further clustering across the full range of scores. This problem is clear in Table 26, where in some cases a raw score range of only 0-15 is being fitted to a standardised score range of 1-19. Ideally, the range of possible raw scores should always be greater than the range of standardised scores, to permit a close fit to the desired distribution on standardisation.

Ten of the measures had *p*-values of less than .05 for the Kolmogorov-Smirnov test, indicating that these samples were significantly different from a normal distribution (α=.05). A visual examination of the distributions revealed that in all cases this was due to clustering at the ceiling scores, as described above.

Due to the nature of the data available, there was no step that could be taken to make the raw data fit a standardised normal distribution more closely. Therefore, there is some question about whether standard scores are the most appropriate method of presenting normative data for these tests. For this reason, percentile norms were also calculated and are presented in Table 28 and Table 29. In fact, these scores were calculated using the same procedure as the standardised scores, with the exception that the final step of matching percentile ranges to standard scores was not undertaken. In this sense, these norms do not differ from those presented in Table 26. However, it can be argued that implicit in standardised scores is the assumption that scores are normally distributed, an assumption that in some cases was inappropriate. Nonetheless, both sets of data are presented, with the understanding that any future use should be with an awareness of the nature of the underlying data, and thus knowledge of the appropriate use of the norms.

Performance of the ED Group

The proceeding sections investigating the psychometric properties of the adapted measures used in this research provided information on their reliability and validity, and presented preliminary normative information on these measures. However, these analyses either report overall sample results, or (in cases where comparison measures are included) analyses based only on the TBI and normative groups. These analyses are important for examining the properties of the adapted measures. Additionally, however, a closer focus on the performance of the 10 individuals from the ED group was required. Were these participants actually able to complete the task requirements? Do the measures tap an appropriate range of abilities? Procedural and practical aspects of these assessments are discussed in *Qualitative Clinical Findings of the Study,* in the next chapter. The following section presents quantitative information to speak to these questions also.

Table 28. Percentile conversion table for attention and memory measures.

	GAT	DF	DB	VPA	VPB	LMA	LMB	FA	FB	FPA	FPB
99			14					46-48	45-48		
97		16						45	44		32
95			13					44		32	
93		15					15				
91								43			
89		14	12	32		15			43		31
87										31	
85								42			
83			11						42		
81								41		30	30
79		13									
77								40	41		
75							14	39			
73	7		10	31						29	29
71											
69		12							40		
67											
65								38			
63						14					
61					8				39	28	28
59			9	30							
57								37			
55		11					13				
53											
51								36	38		27
49										27	
47								35			
45			8	29							
43									37		26
41		10									
39				28						26	
37				27			12	34			25
35						13			36		
33	6			26						25	
31			7								
29		9		25				33		24	24
27									35		
25				24						23	23
23							11				22
21				23	7			32	34	22	
19			6							21	21
17				22		12					20
15				21				31	33	20	
13	5	8		20	6				32		
11				18-19			10	30		19	19
9				16-17				29	31	18	18
7			5	14-15	5	11				17	17
5				12-13			9	28	30	15-16	15-16
3	4		4	11	4	10		26-27	29	14	14
1	≤ 3	≤ 6-7	≤ 3	≤ 10	≤ 3	≤ 9	≤ 8	≤ 25	≤ 28	≤ 13	≤ 13

Note:
Percentiles listed (in bold on left) represent midpoints of 2% ranges.

GAT - Graded Attentional Test; DF - Digits Forwards; DB - Digits Backwards; VPA and VPB - Verbal Paired Associates A and B; LMA and LMB - Logical Memory A and B; FA and FB - Faces A and B; FPA and FPB - Family Pictures A and B. All data is from adapted research measures.

Table 29. Percentile conversion table for other measures.

	AR	BNT	BNT (all cues)	FANL-C Novel	FANL-C Familiar	VOT[a]	MR	MASC Total	MASC Cat.	MASC Pers.
99							26	40	20	≤ 7
97							25			8
95										
93						30		39	21	
91		30					24			
89										9
87										
85							23			
83						29			22	
81										
79	50									
77								38		
75				20		28	22			
73										
71										
69		29	30						23	
67										
65										
63										
61					20	27	21			
59										10
57										
55										
53								37	24	
51										
49						26				
47		28					20			
45	49									
43										
41									25	
39						25	19			
37				19				36		
35									26	
33										
31		27					18			
29			29			24		35	27	
27										
25	48						17		28	
23		26						34	29	
21										
19				18			16		30	11
17	47					23		33	31	
15		25	28						32-33	
13				17	19		15	32	34	
11	46	24						30-31	35-36	
9	45			16		22	14	29	37-39	
7	44	23	27				13	27-28	40-44	
5		22		15		21	10-12	24-26	45	
3	41-43	17-21	26	12-14	17-18	20	9	22-23	46-48	≥ 12
1	≤ 40	≤ 16	≤ 25	≤ 11	≤ 16	≤ 19	≤ 8	≤ 21	≥ 49	

Note:

[a] See Limitations of the Research, Chapter 9, p. 214 for a caution in the use of this measure.

Percentiles listed (in bold on left) represent midpoints of 2% ranges.

AR – Auditory Reception; BNT – Boston Naming Test; FANL-C – Familiar and Novel Language Comprehension Test; VOT – Hooper Visual Organization Test; MR – Matrix Reasoning; MASC – Match and Shift Categories Test. Excluding MR, data are for research versions of measures.

Table 30 presents descriptive statistics on raw scores obtained by the individuals with expressive disabilities. The observed range for each variable is reported, with the possible range of scores that can be obtained on the measure. The table also reports the median, mean and standard deviation for each variable for this group.

One of the concerns was to ensure that adapted measures had a sufficiently low floor that the vast majority of people with expressive disabilities could respond correctly to at least some items. This was considered desirable for two reasons. Firstly, experiencing success in the assessment process can maintain rapport, particularly important where examinees are having to exert considerable effort to respond due to their disabilities.

Secondly, where individuals with expressive disabilities fail all items on a test, a strong hypothesis to explain their performance must be that either they did not understand the test instructions, or cannot reliability use the communication technique being employed. It may be very difficult to establish the extent to which these factors are impacting on performance.

An examination of raw score ranges in Table 30 suggests that for the majority of measures, participants were scoring above the floor scores. However, on a number of measures there were participants who did achieve scores of zero. Nonetheless, median and mean values are in close agreement, and while they place average performance at very low levels, it is clear from the higher scores obtained by participants in this group that the presence of expressive disabilities did not automatically exclude high performance on most of these tests.

An examination of this table suggests that participants had the most difficulty with a number of the memory measures—Digits Backwards, Verbal Paired Associates A and B, Faces A and B, and Family Pictures A and B. Note for Faces A and B that while the range of possible scores on this measure extends down to 0, a score of 24 represents chance levels.

Table 30. Raw score descriptive statistics for the ED group.

Variable	Minimum	Maximum	Possible Range	Median	Mean	Standard Deviation
GSH Orientation	7	10	0-10	9.5	9.20	1.03
Graded Attentional Test	0	6	0-7	2	2.30	2.06
Digits Forwards	1	11	0-16	6	5.60	3.06
Digits Backwards	0	6	0-14	3	2.80	1.87
Verbal Paired Associates A	0	9	0-32	4.5	4.60	2.46
Verbal Paired Associates B	0	2	0-8	1	1.10	0.88
VPA Recognition	11	24	0-24	17	16.70	4.64
Logical Memory A	7	13	0-15	10	9.80	1.81
Logical Memory B	6	12	0-15	9	8.90	1.73
Faces A	19	29	0-48	24.5	24.40	2.68
Faces B	21	32	0-48	25	25.70	3.92
Family Pictures A	4	11	0-48	7	6.70	2.21
Family Pictures B	5	12	0-48	9	8.60	2.50
Auditory Reception	30	50	0-50	41	41.10	6.35
Boston Naming Test	8	23	0-30	19	17.63	5.61
Boston (all cues)	14	27	0-30	25	22.38	5.53
FANL-C Novel	3	18	0-20	9	8.70	4.57
FANL-C Familiar	6	18	0-20	13	12.60	4.03
Hooper VOT	0	20	0-20	9	9.63	5.81
Matrix Reasoning	2	13	0-26	6	6.10	3.90
MASC Total Correct	11	37	0-40	26.5	25.38	8.78
MASC Categories	21	60	0-60	39	41.50	13.53
MASC Perseveration	5	12	0-40	10.5	9.50	3.11

Boxplots were prepared in order to provide a graphical illustration of the performance of each group on these measures. The boxplots display standardised scores on each of the research measures, by group, and are presented in Figure A2 through Figure A22 (Appendix K, p. 385). An example boxplot is presented in Figure 7. Whiskers of the plots represent the range of data, excluding outliers and extreme values, which are indicated individually on the plots. Outliers were defined as values 1.5 to 3 box-lengths (i.e., the inter-quartile range) from the box; extreme values were those greater than 3 box-lengths from the box. While for the purposes of the boxplots these values were labelled as outliers or extreme values, for reasons discussed in the early part of this chapter these were not considered invalid data, and were included in all analyses.

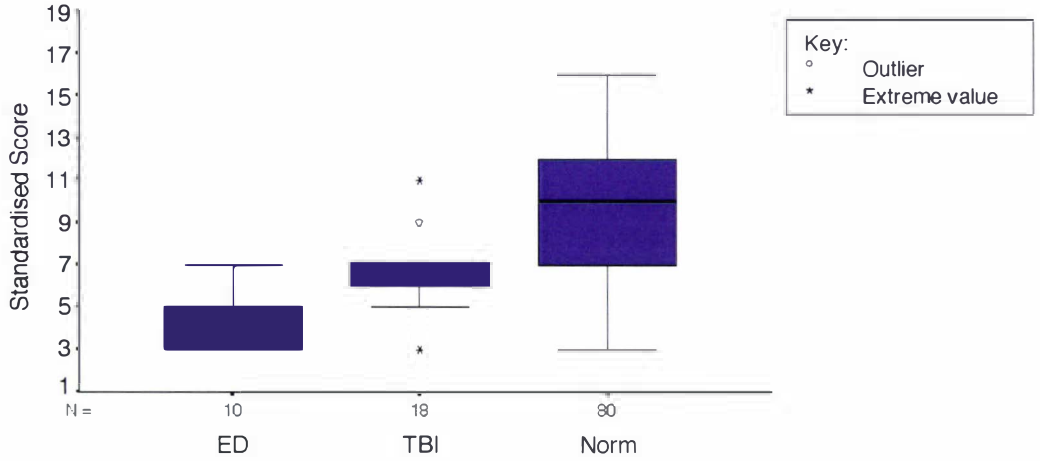


Figure 7. Example boxplot of Digits Backwards standardised scores, by group.

A comparison of the conversion criteria in Table 26 with the scores obtained by ED group participants (Table 30) indicates that a considerable part of the range of scores obtained by participants in this group were collapsed into a standard score of 3 for many variables. This is apparent in the boxplots for this group. That many ED group participants' scores receive a standard score of 3 is due both to statistical limitations on the range of standardised scores (imposed by the size of the normative sample; see Appendix J), and to the fact that participants in the normative group simply did not score at the lower end of the ED group scores.

The distribution of scores on each of these measures, by both the ED group members and the full sample, is evaluated in more detail in the following chapter.

CHAPTER 9: DISCUSSION

Overview of the Discussion

Following on from the previous chapter, the discussion commences with an examination of the quantitative findings from the clinical phase. The extent to which the measures included in the research protocol are useful psychometric instruments for the assessment of individuals with expressive disabilities is evaluated. Following this, the *qualitative* findings of the clinical phase are discussed. Rather than formally collated data, these observations stem from the use of the protocol measures with both the severely disabled participants, as well as the comparison groups. It is considered that such findings are equally as relevant as the quantitative findings, when use of these instruments in a clinical, rather than research, context is being considered. Limitations of the research are then outlined, and the effects that these may have had on the results are examined. Finally, the research questions posed in Chapter 1 are restated and each is briefly discussed, drawing together the pertinent information from the overall research programme.

Quantitative Findings of the Clinical Phase

Test Administration and Delayed Recall Times

Test administration times in the ED group supported the results of the earlier studies reported in this dissertation (Chapter 3 and Chapter 5), and clinical anecdote which suggested that assessment takes considerably longer with this group than other participants. The research measures constituted a little over half of the overall protocol (at 57% of assessment time for the TBI group and 59% for the normative group), but for initial assessments ED group participants took nearly as long to complete these research measures as other brain-injured individuals took to complete the whole protocol. Therefore, the ED group took about 65% longer to complete the research measures. Delay recall intervals also varied significantly between groups for some memory measures. Since the material presented in these intervals was (generally) kept constant[35], the actual time spent (distraction) was dependent on how long participants took to complete the tasks. ED group participants, having more difficulty, accordingly used more time.

---

[35] See *Qualitative Findings of the Clinical Phase*, below.

The problem of what to include in delay intervals is analogous to the need for a placebo control in psychotherapy outcome research. A delay interval with no testing is not normally selected, presumably because this would allow participants to rehearse material. If it was desired to keep intervals constant, a distracter task that did not constitute a measure in the assessment could be used, allowing the task to be discontinued when the interval is up. However, issues of varying fatigue between groups make even time a relative concept. Again, the issue can only be resolved by weighing up the costs and benefits of the selected method for answering particular research (or clinical) questions.

It must be noted that although it was important to record administration time to quantify one of the ways in which task requirements may have differed between groups, the inclusion of these variables in the other analyses would have been complicated. Administration time is related to both ability and final scores. Many individuals with brain injuries exhibit increased response latency, presumably due at least partly to slowed information processing; it is likely that test completion time would be correlated with scores in many cases. Therefore, controlling for administration and delay times as covariates might eliminate variance that is actually attributable to the construct of interest. Indeed, this is recognised by test publishers, who include response times as part of the dependent variable, giving bonus points for quick performance on some tests (e.g., Block Design).

Future researchers considering recording administration times should be aware that the recording, processing, verifying and analysis time required to use such data is substantial. For this reason, blanket recording across an entire protocol (as was done in the current study) could not be recommended as routine practice. However, collecting them may be warranted where the data were required for a specific purpose, such as comparing delayed recall intervals between groups.

Reliability

Internal Consistency

A coefficient above .70 is often cited as a benchmark for acceptable reliability, although such decisions must be based on the nature of the underlying construct

(Franzen, 1989). As the measures included in this research study were designed to measure cognitive functions, considered relatively stable, a high level of reliability is desirable. In this case, the majority of tests in the protocol had internal consistency coefficients of over .85. Considering the fact that many measures were shortened from their standard administrations this is a very acceptable level. Faces A and B were found to have somewhat lower alpha coefficients (.76 and .78) but this test was administered almost in its standard form, and these coefficients are in fact marginally higher than that reported in the WMS-III standardisation study (.74; Wechsler, 1997a). Logical Memory A and B had low alpha coefficients, which may have been partly due to the fact that they were among the shortest tests in the protocol. The finding is nonetheless surprising given that the material appears appropriate and there was a reasonable range of observed scores on the test. Of the other particularly short measures, the Graded Attentional Test had acceptable internal consistency (.79) while the Good Samaritan Hospital Orientation test had an alpha coefficient of .42. However, for this latter test, variability was considerably constrained by ceiling effects, limiting the possible alpha score range.

Comparing the reliability of the adapted and standard measures in the protocol provided a benchmark that controlled for the effects that particular groups and individuals in the research, and unequal cell sizes, may have had on reliability. Of the five measures that had direct analogues for which internal consistency coefficients could be calculated, four showed similar levels of reliability, or suggested the superiority of the research measures. In contrast, Logical Memory had substantially lower internal consistency than its comparison measure. The first four tests were all based directly on their comparison measures, and provide strong evidence that the adaptation to a response card administration does not compromise internal consistency of neuropsychological tests. The low scores observed for Logical Memory would therefore appear to be related to the particular adaptations made to that measure.

Test-Retest Reliability

Acceptable levels of test-retest reliability depend on the nature of the construct being examined, and in this research with short re-test intervals, little variation in the actual abilities of participants was expected. It was hypothesised that one symptom of adaptation to a multiple-choice format would be that learning effects over repeated testing would be increased on many measures. However, provided that such a change is uniform across individuals, Pearson's *r* would not be affected, and thus the results would not be biased by these learning effects.

In general, test-retest reliability was acceptable, with a third of the measures having coefficients of .85 or greater, and the majority over .70. Logical Memory and the Good Samaritan Hospital Orientation test appeared to be less stable, but not low enough to be completely rejected. On the other hand, the three indexes from the Match and Shift Categories test had low and in one case essentially zero retest reliability, raising questions about the construction of this measure. However, this measure (like the Wisconsin Card Sorting Test) relies partly on examinees being unaware of the exact task requirements. Therefore, lower reliability scores may reflect changed knowledge of task requirements on re-test, rather than instability in the trait being examined. This remains a significant flaw in the construction of the Match and Shift Categories Test, but does not necessarily render the test unusable.

Validity

Concurrent Validity

A key source of evidence for the validity of the adapted measures in this protocol was the extent to which they measured similar constructs to their comparison measures. With the exception of the Match and Shift Categories test, all measures were statistically significantly correlated with their comparison measures. Researchers would not usually expect correlations close to 1.0 between alternative forms (Cicchetti, 1994). In this research, however, very high correlations (e.g., over .90) would be required to conclude that adaptation to response cards had not altered the construct being measured at all. In fact, among the significantly correlated tests, only the Boston Naming Test correlations met this criterion. Other correlations ranged from .80 down to .45, with this last being between the two halves of the Hooper Visual Organization Test, indicating they shared only 20% of their variance.

However, correlations between measures are affected by measurement error, as discussed earlier. When corrected for unreliability, a much higher proportion of shared variance was explained, rising for the Hooper from 20% to 81%, a figure higher than that reported in previous studies of the corrected split-half reliability of this test (see Chapter 7). For the Digit Span measures the corrected correlations were .95 and .98, and best estimate for the correlations between the Boston Naming Test scores was 1.00. In fact, for the Boston Naming Test indexes, calculated values of the correlations corrected for unreliability were greater than 1.00, which might lead to the conclusion that such corrections are in some way themselves unreliable. This statement is both true, and yet misleading to the extent that unreliability is associated with unacceptability. Schmidt and Hunter (1996) discuss that in the population (i.e., with an infinite sample size) the correction for attenuation contains no error. However, with smaller sample sizes, the correction is an estimate, and contains sampling error. Therefore, when the true correlation is 1.0, researchers would observe corrected sample correlations over 1.0 in 50% of studies (Schmidt & Hunter, 1996). The appropriate course of action in this case is to round the corrected correlation down to 1.0, as this is the best estimate of the true relationship (Schmidt & Hunter, 1996). Using these corrected correlations, five measures met the arbitrary criterion (mentioned above) of a correlation of over .90. Other measures, that did not meet this criterion, nonetheless had high correlations. These findings suggest that the adapted measures do indeed continue to measure highly related constructs, if not identical abilities to the original measures.

Construct Validity

Examination of the issue of construct validity through factor analysis provided some limited further support for the suitability of the adaptations. Factor analysis may have provided a clearer picture had the various constructs of interest been less generally correlated. Due to the broad first factor, which accounted for over half of the shared variance in the sample, more subtle interrelationships between measures, and particularly between research and comparison measures, were difficult to detect. However, in two areas, that of language measures and the digit span tests, further support for the construct validity of the adaptations was provided. In a general sense, the results of the factor analysis do support the validity of the adapted measures, since they are clearly related to performance on a wide range of other tests that tap cognitive functioning.

Discriminant Validity

The presence of statistically significant effects of demographic variables on test performance clearly limit the extent to which inferences can be made from analyses of group differences in the sample. While individuals in the sample were not matched on demographic variables, efforts were made to deliberately recruit individuals in the target demographic groups. It is noteworthy that recruiting appropriate samples is not a trivial matter—despite efforts to recruit a balanced normative group, when two thirds of the normative sample had been collected, there was a three-to-one ratio of females to males in the sample. A period of months followed where females were not seen and advertising specifically targeted males in order to attempt to address this imbalance. However, gender balance could not be achieved at all in the other groups. Similar difficulties were found with recruiting a wide range of educational backgrounds. Although relatively few participants in the normative group had university level education (due to targeted advertising), many more in this group had tertiary certificate and diploma qualifications than in the other groups.

Despite these concerns, group differences on the research measures were significant for all tests except MASC Perseveration, and effect sizes indicated these differences were large. Planned comparisons of group differences partially supported the hypothesised sensitivity of the majority of measures to differences between both the ED and TBI groups, and the TBI and normative groups. For two of the language measures (Auditory Reception and the FANL-C Novel scale), and the Graded Attentional Test, the TBI and normative groups did not differ, but significant differences were found between the ED and TBI groups. That individuals with expressive disabilities had lower scores on these language measures is not surprising since they have an identified disruption in language–communication systems. However, the finding that there were TBI–Norm group differences on some language measures but not others suggests that these tests do measure different aspects of language functioning, and that a broad range of language tests such as this should therefore be considered in any 'comprehensive' assessment.

Summary of Psychometric Properties of the Research Measures

Generally, the research measures were found to have acceptable levels of reliability and validity, and to be suitable measures of the intended domains of cognitive functioning. As a group they also clearly demonstrate that measures adapted to response card administrations can maintain appropriate levels of reliability and validity. However, future adaptations still require research on a test-by-test basis.

While overall results were positive, a number of measures in the protocol appeared to be less successful. Some of the difficulties found with the Match and Shift Categories test may be due to the methodology employed in the current study, since the initial administration is thought to have altered performance substantially on both the comparison measure and on subsequent re-testing with the Match and Shift Categories Test itself. Nonetheless, the evidence does not support the use of this measure without further research to improve it.

While the adapted Logical Memory test appeared to have some validity, results were not as robust as for some other measures. This may be partly due to the fact that the adaptations to this measure resulted in one of the most radical changes in content, in administration and reduction in difficulty, and this test was the most different to its comparison measure. Thus, the comparison may not have been entirely reasonable. Nonetheless, a more comprehensive test along similar lines could productively be examined in the future (e.g., more items, multiple-choice rather than Yes/No, less obvious answers).

There has been only one study similar to the current project reported in the literature, conducted by Berninger et al. (1988), against which the current findings could be compared. That study adapted all measures to a multiple-choice format, where examinee choice was indicated with a series of yes/no responses. In their two comparable analyses (test-retest reliability, and concurrent validity with standard versions of measures), coefficients of reliability and validity were similar to those for the measures examined in the current project.

No other reported studies have conducted research along the lines of the current project. Therefore, further comparison of the current results with a wider body of research is not possible. It is to be hoped that in the future further research will investigate the assessment of individuals with this level of disability, however.

Normative Data

The normative information reported in the previous chapter provides the basis for cautious interpretation of future clinical use of these adapted measures. However, these tests fall well short of the criteria for being standardised (American Psychological Association, 1985), both in terms of the evidence for the psychometric properties of the measures and the normative information ultimately produced, but also the process by which these tests were examined (see Mapou, 1988, and Mitrushina, Boone, and D'Elia, 1999, for two outlines of best practice in this regard). This is not specifically a flaw in the research, since the aims of the research were to examine assessment options for clinicians working with this population, rather than produce specific standardised tests.

The normative information reported is based on a relatively small sample size by current neuropsychological standards. However, while some existing published tests are suitable for use with people with expressive disabilities, this study reports one of the most substantial known attempts to develop a comprehensive assessment protocol for use with this group. As such, the current normative sample is of practical and clinical significance. Due to the limited variability in the normative sample above the 50th percentile for most tests, performance at or above the mean is probably not interpretable other than as 'normal'. In contrast, low performances, particularly using the criterion of a standardised score of three or four, indicates a strong likelihood that the individual is having significant difficulty in that domain of functioning.

As a result of limitations in the calculation of standardised scores, discussed in the previous chapter, some users might consider the percentile norms to be more appropriate for reporting assessments conducted using these instruments. Despite this, the value of standardised scores in the comparison of performance across a range of tests warrants their inclusion here. Additionally, Rust and Golombok (1989) noted that standardised scores such as this are "sufficiently imprecise not to be misleading" (p. 84), while percentiles may imply an unrealistic level of specificity in the tests. That is, a 0-100 scale implies a fine gradation of scores that in many cases is not present in the underlying data. Again, caution would have to be exercised in the use of the reported norms in either form. Despite this warning, they represent the next step up from statistical procedures such as simply differentiating scores from chance performance (e.g., Sinnett & Holen, 1995).

Performance of the Expressive Disabilities Group

In addition to acceptable psychometric properties in the overall sample, to be useful instruments they must provide sensible information about the performance of individuals with expressive disabilities. It has been demonstrated that participants in the group were capable of making the basic responses required to complete the measures. The measures also were reasonably sensitive to scores obtained by the bottom 50 percent of the general population, and possible score ranges extend down below scores obtained by the normative group on most measures. However, the measures would be of marginal utility if all individuals with expressive disabilities and cognitive difficulties score zero on every scale[36]. Figure A2 to Figure A22 (Appendix K, p. 385) therefore presented boxplots of standardised scores by group on each of the research measures.

In light of the large differences found between groups in the earlier analyses, it was to be expected that participants in the ED group would generally score in a relatively low band of standardised scores, regardless of the actual variability within the group. However, where scores were uniformly low the question was raised as to whether performance was unduly influenced by physical disability factors in addition to

---

[36] Such a group of measures might still differentiate individuals with brain injuries from those with extremely severe physical disabilities largely in absence of cognitive difficulty.

cognitive demands. From an examination of the boxplots, measures of particular concern were Verbal Paired Associates A and B, Faces A, Family Pictures A and B, and the Hooper Visual Organization Test.

An examination of raw scores obtained by the ED group (Table 30, p. 194) provided more specific information about these measures, and appears to bear out some of the concerns. Performance on both Verbal Paired Associates A and B and Faces A is essentially at chance levels on average for this group. It may be that for this group the attentional demands of these measures become more important factors affecting performance than the 'memory' demands. Low scores on the Graded Attentional Test would tend to support the assertion that this was an area of difficulty. Performance on the Hooper Visual Organization Test also appeared from standardised score distributions to be an area of possible concern, but the raw scores show an appropriate range, median and mean, and do not appear to suggest a significant floor effect.

Overall, while scores were substantially lower in this group, these adapted measures appear to have a sufficiently low floor to examine the level of functioning observed. One issue raised by these results is the fact that the importance of various peripheral cognitive abilities involved in completing a task (i.e., abilities other than the construct of interest) may change from group to group. Specifically, tests that are sensitive to fatigue or attention deficits may be particularly affected. Similarly, while it was intended to eliminate timing from measures in the protocol, a number of tests contained a paced presentation of material (e.g., a story is read once at a certain speed; Faces photographs are presented for two seconds each) requiring quick orienting and attention. The examination of tests where pacing can be controlled by the examinee would be of use is exploring these issues further.

Qualitative Findings of the Clinical Phase[37]

During the course of administering the protocol measures to individuals in each of the groups, a number of qualitative observations were made which were pertinent to

---

[37] An early version of this discussion was presented in Babbage and Leathem (1999, September). See Appendix A.

both the use of the measures in this protocol, and the future development of similar measures for use with this group. Firstly, planning for short sessions is extremely important. The initial aim was to be able to have a number of short assessment sessions when designing the protocol for this research, and the concept of 'short' probably equated to 30-minute sessions. On the basis of the experience of conducting the assessments in this research, a protocol that is divided into 10-minute units would seem more suitable. This would cause difficulties for delayed recall trials on memory measures, but the performance of some individuals in this research suggested that it might be more appropriate to establish whether a person has any immediate recall before attempting delayed recall procedures. Measures such as Faces, which have a relatively large number of items, can be frustrating for examinees that strongly believe they cannot remember anything. However, administration of such measures *is* useful, since above-chance performance may be observed in the absence of reported recall.

One goal in the adaptation of the measures used in this research was to maintain as much as possible the original properties of the test, while converting the response required to a response card or multiple-choice format. It was believed that where responses that are more complex could be made through this medium the assessment would provide a clearer picture of functioning, as examinees would be less constrained by the task requirements.

However, it became clear during the data collection that in some ways *less* was *more*. That is, where less information was collected, and participants had fewer choices to select from, data that are more useful might be obtained. The rationale for this conclusion is that where responses were more constrained, there were fewer sources of error. For example, a comparison could be made between the two measures of non-verbal memory, Faces and Family Pictures. For Faces, participants answered yes or no for each item; an answer is always given, and if given it was the most complete answer possible. In contrast, for Family Pictures, there were levels of possible performance. In scoring descriptions of the family members in various scenes, Family Pictures distinguishes 1-point and 2-point responses, based on the amount of information given. A significant number of participants in this study, in both the TBI and normative groups, tended to spell out a single word on the alphabet board (e.g., "eating"), and then follow this (verbally) with "off a plate, and there is a check tablecloth, and he has a fork in his hand about to go to his mouth, and..." This was done in a context where each time participants spoke (during the research measures administration) they were politely reminded that they needed to give their answers via the response cards. The use of response cards was rigidly maintained. Despite this, participants continued to make these types of statements. Clearly, these participants recalled more than they were communicating on the alphabet board[38], and yet decided not to communicate all of the information available to them. That they continued to verbalise responses may have been due to a range of factors, including (perhaps ironically, since they were not given credit for spoken responses) a wish to communicate as much as possible. Related to this point, a number of participants in the normative group communicated exhaustive detail via the alphabet boards, indicating that the task requirements were unclear or at least ambiguous in suggesting the amount of detail that was required.

If this behaviour (of not providing all known information when using an alphabet board) was observed in both the TBI and normative groups, could it also be occurring in the ED group? A number of views could be taken on this issue. Various factors suggest that severely disabled participants might be more likely to communicate all available information. Firstly, they have no other mechanism to use (such as 'blurting out' the answer) to surreptitiously indicate to the examiner that they know more than they are letting on. Secondly, individuals with expressive disabilities are generally highly motivated to use the alphabet board or LightWriter, as it is their primary source of communication with people, including the examiner. This obviously does not hold true for the comparison groups. Additionally, individuals with expressive disabilities will generally be very familiar with this style of communication technology and therefore may be more comfortable using it.

On the other hand, it is likely that examinees in this group have been strongly reinforced over a long period for communicating in a clipped, abbreviated fashion.

---

[38] In all cases, participants were scored for material provided via the alphabet board.

Having experienced frustration from people who will not wait for a full paragraph (or even a sentence) to be spelled out, people with expressive disabilities may abbreviate their messages in order to ensure they keep the attention of the recipient long enough for communication to occur. The extent to which this could be modified, by providing encouragement to extend communications in the testing session, was not investigated in the current study. Certainly, if using measures such as Family Pictures with this group in the future, the administration should be altered so that probes are given to encourage all participants to communicate the full amount of information they have available. (Such a change, however, would require normative information specific to that procedure.)

People with expressive disabilities may be more likely to exhibit fatigue due to the communication process and may shorten their communications as a result. Fatigue is a major issue with this group. Subtests take longer to administer, yet these participants need shorter sessions. Some tests cannot be administered within the 'fatigue window', resulting in compromised results whether the measure is divided into parts or testing is continued into a period of sub-optimal performance. The best clinical decision to make here would depend very much on the referral questions to be answered. For example, if the question relates partly to a person's ability to sustain a certain activity, fatigued performance may be an important part of such an assessment. However, the fact that scores were obtained under such conditions should be clearly stated in any reports, as is standard best practice.

One of the effects of fatigue in the current research was the requirement to vary the test protocol administration order for individuals in the ED group (and occasionally the TBI group). It was difficult to judge *a priori* the speed with which each participant in the ED group would fatigue, despite investigating this issue with caregivers and family members (as well as directly discussing fatigue with participants). Consequently, it was not uncommon for test administration order to have to be altered in the midst of an assessment, due to a person's rapidly reducing ability to participate. (Despite at times being very keen to continue.) As mentioned above, the key difficulty was the administration of delayed recall trials for memory measures, which were contained in (and indeed delineated) each section of the protocol. Each quarter of the full protocol contained two inter-linked memory measures (see Table 13, p. 120). As ED group participants completed only the research measures, this arrangement provided only one break, in the middle of the protocol, which was often grossly inadequate. Frequently, the memory measures were therefore un-linked, with the delayed recall trial of the first measure performed *before* the initial trial of the next measure, dividing the protocol into quarters. This provided the opportunity for three breaks in the assessment. In a number of cases this was not sufficient, and while memory and delayed recall trials were always performed within the one session, other tests were re-arranged on occasions into blocks of appropriate length. These alterations could have affected the results. However, reducing the amount of distracter material before a delayed recall trial could only improve the ED group individual's performance, and thus such changes are conservative relative to the hypotheses examined.

Differences in task requirements and administration under conditions of relatively low performance must also be carefully considered. For example, Verbal Paired Associates, both in its standard format and in the adapted administration used in this research, provides feedback of the correct response when an error is made. Individuals experiencing only moderate difficulty on this task received this feedback only infrequently, and could often use this information productively to aid their performance. However, for a number of participants with brain injuries, this feedback did not seem to assist them in producing more correct answers, and as a result they received a great deal of correction. This appeared to have a tendency to heighten their awareness of their difficulties and produce negative emotional responses which were both uncomfortable for the participant and which may have further reduced performance. Such feedback following incorrect answers seems particularly inappropriate in the context of multiple-choice answers, where the answer is literally 'staring them in the face'.

Another observation was that perseveration seemed to play a significant role in the responses of some participants, who repeatedly selected the same response. Perseverative tendencies may tend to be enhanced in testing situations which use multiple-choice answers, since the self-generation of responses is not required to fulfil the task requirements, and similar responses are repeated from trial to trial. An investigation of these issues is warranted.

Two participants in the ED group had particular difficulty with measures that required the use of alphabet boards, despite being able to handle the physical aspects of using the board. In these two cases, some subtests were discontinued. It may be that these difficulties were related to an expressive aphasia that affected more than vocal communication. Certainly, other participants in this group who were able to use the alphabet board 'successfully' in terms of producing responses which could be scored still performed at low levels, indicating language difficulties.

Two potential participants in the ED group were excluded due to the effects of massive fatigue making such a comprehensive range of assessment impracticable. With the assistance of rehabilitation service staff (for postural support, a familiar person to guide communication efforts, and so on) it was demonstrated that these participants were able to make reliable responses using the response cards, although this was not a usual mode of communication for them. While these efforts did not lead to a formal assessment, useful clinical information was gleaned from the attempts. With one individual, one of the rehabilitation staff was that person's speech language therapist. The attempts at communication using the mechanisms in this research provided some further information that could inform the ongoing therapy work. Such assessments could potentially uncover previously unknown areas of both deficits and strengths. Therefore, the rehabilitation effects of such an assessment attempt may work in ways other than solely through the formal assessment report and recommendations.

Suggestions for Clinical Assessment

Taking into account reliability and validity information and specific findings with the ED group, there did not seem to be any clear trends regarding the most appropriate kinds of response card adaptations to make. Rather, content of measures and administration factors such as those discussed above appeared to be the important factors relating to the ability of participants to complete the tasks. Future research could fruitfully examine the issue of whether almost identical measures using different response formats differ significantly in terms of psychometric properties, but due to the significant differences in content between measures in the current research, this fine-grained analysis was not possible.

In general, clinicians assessing individuals with expressive disabilities should utilise published tests with large, applicable normative samples. Where such tests are not available, partially standardised tests such as the ones described in this research may be suitable. Finally, where neither of these alternatives are possible, adaptation to a multiple-choice format would allow analysis relative to chance performance (see Sinnett & Holen, 1995). Such questions may seem basic in some contexts, but a recently published case regarding the maintenance of extreme life support measures highlights the importance of such assessment methodologies. In this case, two assessments by separate neuropsychology services were independently requested by the court to resolve a difficult decision (reported in McMillan, 1997, and Shiel & Wilson, 1998). Other assessment methods may also be possible in some contexts with such individuals. For example, recent research (mentioned at the beginning of this dissertation) investigated the use of event related potentials as the dependent measure in cognitive testing for individuals incapable of making outwardly visible responses (Connolly, Mate-Kole, & Joyce, 1999).

Limitations of the Research

Limitations of the Design

As previously discussed in some detail, the confound of demographic variables with group status complicates the interpretation of some of the results of this research. Perhaps a more fundamental design flaw was the absence of a group with expressive disabilities in the absence of neuropathology (labelled the physical disability, or PD, group, in Chapter 7). However, while an ideal design would include such a group, efforts to recruit people who fitted these criteria were unsuccessful.

However, even if a suitable PD group had been found, MANOVA analysis examining the main effects and interactions of physical disability and brain injury as dichotomised variables might not have been appropriate. In the current research it was argued that, as extremes of their continua, these two variables could be treated as dichotomies. It was possible to argue that there was an absence of physical disabilities in both the TBI and normative groups. However, had a disabled, non-brain-injured PD group with expressive disabilities been recruited, it would have been a substantial hurdle to demonstrate that the level of expressive disability in the ED and PD groups had been the same. It would also have been difficult to demonstrate that the severity of traumatic brain injury in the ED and TBI groups was the same. These two conditions would have to be met in order to conduct the originally planned analyses. If these conditions could not be met, the results of the MANOVA would have been questionable and could have been an artefact of the differences between these groups. For example, if there was a significant interaction effect, it might have just been an additive effect with expressive disabilities and traumatic brain injuries being more severely represented in the ED group than their 'matched' comparison groups. Therefore, the current design, which did not require traumatic brain injury and expressive disability severity to be quantified exactly, was appropriate as it was more robust to differences in injury severity between the groups.

Ideally, a larger normative sample would be recruited, and the adapted measures in this research would be fully standardised. However, while this is a limitation of the research, a larger study would have been impractical in the current context.

Memory Test Delay Intervals

For memory tests with delayed recall or recognition trials the decision was taken to keep delay interval content the same across groups. It was recognised that this would likely result in varying delay intervals between individuals and the results supported this expectation. The broad range of delay intervals observed is a limitation of the research, confounding interpretation of delayed recall and recognition performance. As discussed, the extent to which such variation occurs in other studies is not clear, since few researchers report the range of delay intervals that occurred. Although not practical in the current research, a robust approach for future research would be to control delay interval by using a timed distracter task of the appropriate length.

Hooper Visual Organization Test

Original Hooper Visual Organization Test materials had been re-bound into a booklet also containing other measures, for the purposes of this research. During the course of the research the first stimulus booklet required re-binding due to heavy use; this coincided with the accumulation of a second set of the assessment materials, including re-binding of the Hooper Visual Organization Test materials for this second kit.

Unfortunately, staff at the service which undertook the binding re-bound the Hooper Visual Organization Test stimuli upside down and due to their ambiguous and already rotated nature this error was not discovered until late in the project. Due to the order the assessments were conducted, approximately half the participants in the ED and TBI groups received the original orientation, and half the 180° rotated orientation. However, only 10% of the normative group participants received the original orientation. In all cases, both the research and comparison measures were rotated. Mean scores for the original and rotated orientations were compared for each group, on both the research and the comparison measure. No significant differences were found between the two versions, in any group, and so data were collapsed over the two rotated versions in the analyses. Nonetheless, the results should be interpreted with this unusual presentation in mind.

Cultural Bias in the Research Protocol

With the exception of the two tests developed for this research, all the measures used were created and originally normed outside of New Zealand. While in recent years there have been efforts by test publishers to reduce culture-specific material in their scales (see, for example, Wechsler, 1997a), significant amounts of such bias still exist. Some test stimuli depict animals specific to geographical areas (e.g., Racoon, Verbal Paired Associates; Beaver, Boston Naming Test). Tests which included pictures of people either depicted North American Europeans (e.g., Family Pictures) or a cross-section of the USA population (Faces includes some people of Latin-American, African-American and Asian descent, but does not include Maori or Pacific Islanders as would be most appropriate for the New Zealand context). As the decision was made to develop this research from the basis of existing instruments, these factors could not be controlled. However, future attention to them is warranted.

Another interesting observation is that there were a number of stimuli which appeared in more than one of the protocol measures (e.g., Sphinx appears in the Boston Naming Test and Auditory Reception; Scissors in the Hooper VOT and Boston Naming Test comparison measure; House in the Boston Naming Test comparison measure and Auditory Verbal Learning Test). There is the potential for a priming effect here which could influence outcome and this issue deserves attention in the future when combining tests from different sources.

Gender Bias in the Research Protocol

Participants and the research assistants identified one instance of gender bias or gender-role stereotype in the research protocol. The item that was questioned was from Auditory Reception, "Do brides dream?". A small number of participants held strong views about the inappropriateness of this item. For this reason, it should probably be excluded. From a purely technical standpoint, as people in the general population do not automatically agree on one answer to this question, it is an inappropriate item since it appears to tap political and gender-role beliefs more than verbal comprehension. It is suggested that this item should be excluded from future uses of this scale, and scores pro-rated to obtain a total score.

Findings of the Research Summarised

Based on the findings of the clinical phase, and the preceding studies, this section attempts to answer each of the research questions that were posed in Chapter 1. The research questions are restated, and the key findings relevant to each question are briefly reviewed. Specific results were presented in the preceding chapters.

For individuals with expressive disabilities,

- Are there any domains of cognitive functioning in which existing psychometric assessment measures are unsuitable?

A review of the files of 271 clients seen over a 9 year period at the Psychology Clinic, Massey University, revealed that for individuals with expressive disabilities, many domains of functioning required the use of adapted or new assessment procedures. There was no individual in this group who could be assessed solely with standard measures.

A survey of neuropsychological practitioners revealed that regardless of theoretical orientation to test selection and administration, participants stated that their usual testing practices would have to be considerably modified, with only 36% of their primary measures, and 47% of their secondary measures, considered suitable for use with the individual in a presented case vignette.

In selecting a protocol of measures for use in the clinical phase, a number of measures were suitable with no adaptations (Good Samaritan Hospital Orientation Test, the Familiar and Novel Language Comprehension Test, Matrix Reasoning) or with minor adaptations (Auditory Reception, Faces). However, for many domains existing psychometric instruments were unsuitable and required adaptation.

- Can existing psychometric measures be adapted to be suitable, or new assessment measures created?

In some cases existing psychometric measures were adapted to be suitable for use with people with expressive disabilities, by changing the response requirements so that participants did not need to speak, write, draw, or manipulate test materials, and removing all timed components for responses. Instead, the adapted measures required participants to point to answers on response cards, or to spell out answers using an alphabet board or LightWriter. A number of existing measures were adapted to this format, and two new measures were developed, based on principles from previously published tests.

The results of the clinical phase indicated that most of these adapted tests had reasonable psychometric properties, in that they were reliable measures, were related to the constructs of interest, and could be shown to be sensitive to disruptions of cognitive functioning due to brain injury.

- Are there any domains for which no assessment is possible?

While adapted measures were developed for the assessment of a range of domains, no tests could be selected, adapted, or created, to measure Construction, Information Processing Speed, or Motor Ability in individuals with expressive disabilities.

- What are the psychometric properties of the adapted and new tests, and do they meet the minimum criteria for acceptable measures?

Minimum criteria for what are acceptable measures are not fixed, but relative to the purpose of test, underlying stability of the construct, and a range of other issues. However, based on the initial findings most of the adapted measures used in the clinical phase of this research appear to be useful tests with acceptable levels of reliability and validity. To meet the full criteria for acceptable standardised measures, further research establishing their psychometric properties and providing more comprehensive normative information would be required. However, in light of the paucity of information available on conducting assessments with individuals with this level of disability, preliminary normative information was provided based on the normative group.

While overall results were positive, and a number of measures were clearly demonstrated to be useful, questions remain about others. Principal among these are questions about the validity of the Match and Shift Categories test, one of the two measures created for this research. However, due to the particular task requirements the current research may have been an inadequate test of the properties of this measure. Further research is required before this measure could be recommended for use.

- Are the psychometric properties of adapted measures different from the properties of the measures on which they were based?

On the basis of the current findings, it can be stated with reasonable confidence that the majority of the measure assess aspects of the same constructs, or assess similar constructs, to the original measures. In some cases there are clear theoretical reasons for believing the adapted measures tap different functions—for example, where a free-recall memory measure was converted to a recognition format (e.g., Logical

Memory). In other cases, changes may be more subtle, with facets of the response potentially influencing the cognitive systems required to meet the task demands (e.g., Digits Backwards on a number board may promote use of a visuospatial working memory system, if such a cognitive structure exists, in tandem with the normal methods of processing this information). Further research focussing closely on each adaptation would be required to clarify these issues.

- What information, if any, is lost though changes to measures subtly affecting the underlying construct they examine?

It is difficult to provide firm information that speaks to this question. In some cases, information was clearly lost on the adapted measures regarding an individual's ability to self-generate answers, as opposed to responding to multiple-choice cards. However, some alphabet board adaptations permit this free level of responding. Using the current adaptations, the testing environment is even more rigidly structured than a standard neuropsychological assessment, and it is possible that qualitative information about domains such as executive functions would be more difficult to observe on these tests, as examinees are not given as much latitude in the responses they can express.

- What further research could productively be conducted in this area?

Throughout preceding sections, specific aspects of the research that would benefit from further scrutiny were identified. A number of broader research questions also were raised from this research, in terms of the neuropsychological assessment of individuals with expressive disabilities and with other disabilities. One of the objectives of this programme was to evaluate the extent to which, in hindsight, the research questions asked were the right questions to ask. This research has produced useful information on the assessment of individuals with expressive disabilities, suggesting that right questions were asked. However, a number of other questions could have been asked, and these are outlined below.

Firstly, while the focus of this project was solely on the assessment process, and not the outcome of assessment or its use, future research could examine the specific questions which would lead a person with expressive disabilities to be referred for assessment. To what extent can the measures available provide answers to these questions?

The current research focussed only on cognitive functioning, but assessment of individuals with expressive disabilities would ideally include personal and interpersonal functioning assessment as well. Psychometric assessment of these domains may or may not be appropriate in this population; it may be that interview-based assessment would be more suitable.

While this research focussed on traumatic brain injury, it was hoped that the findings would be applicable to other populations, such as people with cerebral palsy or individuals who have experienced strokes and have aphasias. Studies investigating the use of this type of adaptation in these groups (and others) would be useful both to inform the literature on those populations and to provide further evidence indicating the psychometric properties of tests using these response methods.

The current project was placed solely within classical psychometric theory, which remains the dominant force in psychological testing today. However, with the current population, where innovative assessment approaches are required, application of alternative approaches to assessment might prove fruitful. One advantage of these approaches is the possibility that assessment instruments could be flexible enough to provide a low floor and high ceiling, while only requiring a relatively limited number of items to be administered to examinees. Item Response Theory and assessment based on Bayesian theory provide two possible avenues for progress in this direction (see Elwood, 1993, Embretson, 1996, and Jones, 1989). Unfortunately, neuropsychological assessment measures are not traditionally designed along these lines, meaning that a researcher investigating these approaches would be required essentially to begin from scratch.

- What guidelines can be offered to clinicians who are conducting assessments with individuals with severe physical and sensory disabilities, particularly where the adaptations examined in the current research prove unsuitable?

These issues were discussed in greater detail in the earlier section of this chapter. There are a number of techniques available to clinicians where standardised tests are not available. While there are limits to the conclusions that can be taken from such assessments, clinical creativity combined with careful preparation and administration can yield useful information even with extremely disabled individuals.

Summary of Discussion

This chapter began with an examination of the quantitative findings of the research, which overall provided support for the types of adapted assessment examined in this research. Due to a paucity of other studies in this area, there was not a body of literature to relate the current findings to. However, qualitative findings, based on the experience of administering these measures, were also discussed in the hope that these may assist future measure development. A number of limitations of the research were outlined, after which the research questions posed in Chapter 1 were each answered briefly.

Summary of Part III

The clinical phase of this research examined the properties of a research protocol of measures adapted to be suitable for individuals with expressive disabilities. This protocol was developed based on the findings of the earlier studies, presented in Part II. The measures in this protocol were generally found to be reliable and valid neuropsychological assessment instruments and provided support for the types of adaptations trialed in this research.

The following final section revisits the aims and objectives outlined for the research, and evaluates the success of the research against these criteria.

PART IV:

EVALUATION OF THE RESEARCH PROGRAMME

CHAPTER 10: EVALUATION OF THE RESEARCH

Overview

This final section reviews the aims that were outlined at the beginning of this research programme. While structure is similar to the latter part of Chapter 9, that chapter focussed on the information that the studies provided about the research questions themselves. This section is a broader review of the overall aims and objectives of the programme. Objectives are reviewed in turn, and the extent to which each was met is briefly discussed. Based on this, the success of the research programme is evaluated.

Aim 1

That the research is clinically relevant, providing information about realistic assessment options for clinicians working with people with severe physical and sensory disabilities.

This means that the research should:

- Focus on a group of individuals who represent a significant proportion of the disabled population who are likely to present to neuropsychologists for assessment.

In order to ensure that the group selected for the clinical phase of the research was appropriate, the Files Review was conducted to examine the proportions of people with various disabilities who presented to the Psychology Clinic, Massey University, over a nine year period. Despite selecting the most frequently seen group of people with this extent of disability, recruiting participants in this group was difficult due to their relatively small representation in the general population. Nonetheless, as has been discussed in a number of previous sections, such a group represent an important group due to their significant requirements for care and rehabilitation.

- Have a broad-based focus that enables a comprehensive assessment to be made, rather than focussing solely on one domain of cognitive functioning.

This objective was met by a careful analysis of the range of domains that are considered to compromise cognitive functioning, through a review of the literature, and the studies presented in the *Practitioner Survey* and *Journal Review*. Based on this a broad protocol, as comprehensive as possible, was selected for the clinical phase of the research.

- Use methods and equipment that practising clinical psychologists could access. Expensive or unwieldy custom equipment is therefore not suitable. Commercially published psychometric tests that are not commonly available would need to be carefully considered, since if a service sees clients with this level of disability only on rare occasions, they are unlikely to acquire such tests.

In general, this objective was met. The majority of measures used were either commonly used tests, or non-commercial tests that could be readily accessed. A small number of tests were either relatively new (e.g., the Familiar and Novel Language Comprehension Test) or created for this research and were selected for the reasons discussed in Chapter 6.

- Use timeframes that are realistic in clinical practice terms, regarding the length and number of assessment sessions required.

The lengthiest assessment was conducted in around 5 hours of contact time, including consent procedures and short breaks. This compares well with the average time spent in neuropsychological assessment by clinicians in one recent large survey in the USA (8.7 hours including report writing; Sweet, Moberg, & Westergaard, 1996). This timeframe is therefore very realistic.

- Produce a group of adapted psychometric instruments that have been subjected to at least initial psychometric standardisation.

Analyses indicated that the majority of the research measures were of demonstrable use in the assessment of individuals with expressive disabilities. Both standard score norms and percentiles were provided to allow future use of these measures.

- Provide specific suggestions about adapting cognitive assessment measures, for use by clinicians where the measures developed are not suitable in particular cases.

Strategies of making such adaptations were discussed in the previous chapter.

Aim 2

That the research represents a significant contribution to knowledge regarding assessment of individuals with severe physical and sensory disabilities.

This means that the research should:

- Answer the research questions, as proposed in Chapter 1.
- Evaluate the extent to which the research questions were in fact the correct questions to ask, in the light of the findings.
- Propose suitable areas for future research. It was anticipated that this research would raise at least as many questions as it answered.

Each of these were discussed in the previous chapter.

- Be disseminated in a way that informs psychologists, other health professionals, and consumers of health and disability services.

The results of earlier parts of this research were presented at a number of conferences to disseminate the findings among psychologists (Babbage & Leathem, 1997, 1998b, 1999a, 1999b). The results of the research programme will also be published in a number of journal articles, which are currently in the process of being drafted. A presentation was made at a brain injury service consumers' and professionals' conference to bring the research to both consumers and other professionals (Babbage, 1998a).

Aim 3

That the research conducted is ethical, does not cause harm to participants, and minimises any discomfort or distress they might experience either due to, or during, the assessment process.

This means that the research should:

- Be subjected to ethical review by peers and, where appropriate, ethical review committees.

In addition to being reviewed by peers, the clinical phase of this research was scrutinised by the Massey University Human Ethics Committee, and Health Research Council of New Zealand ethics committees in each of the five regions in which data were collected.

- Utilise procedures for obtaining informed consent that are appropriate for the populations being studied, and obtain consent from as wide a group as possible. In the clinical groups, this might in some cases include guardians or caregivers. Under all circumstances, the informed consent would be obtained of participants themselves.

These procedures were followed. See consent procedures, outlined in *Methodology*, Chapter 7.

- Use assessment protocols that are as brief as possible while answering the research questions, and ensure that participants needs are taken into consideration in the administration of the measures. Particular attention should be paid to the needs of participants with disabilities, such as in some cases the effects of increased fatigue and a more frequent requirement for personal care.

While a comprehensive protocol was desired, efforts were made to select short measures and to consider issues of fatigue when designing the protocol. In addition, flexibility in administration of the protocol was allowed, particularly in the severely disabled group.

- Provide opportunity for participants to discontinue assessment at any time, and to not answer any particular question. Additionally, if the participant becomes upset by the assessment process, the examiner would discontinue assessment themselves, following discussion with the participant.

These opportunities were available and at times were participants were specifically asked about the possibility of discontinuation. In no case did participants elect to discontinue on these grounds, however.

Aim 4

That the research is good science, maintaining rigour and objectivity.

This means that the research should:

- Operationally define all concepts in a manner consistent with both research and clinical practice.

This research was based in traditional psychometric theory, the core basis of the field, and in generally used measures or adaptations of measures which have been accepted by the research and clinical community as valid for the constructs of interest. Considerable effort was taken to select both appropriate domains of investigation (see *Practitioner Survey* and *Journal Review*) and measures of these domains (see Chapter 6 and Chapter 7).

- In the clinical phase, take reasonable steps to ensure that research measures are administered in a consistent way across individuals, groups, and examiners.
- Recognise that good science is also flexible, and that the best planned new assessment techniques may prove to be less than ideal when trialed with a clinical population. Nonetheless, the planned administration of measures will be adhered to, to the greatest extent possible. Any deviations from this should be carefully documented, and the implications of such deviation considered.

This consistency was maintained to the greatest extent possible, within the boundaries imposed by the requirements of people with severe disabilities. Alterations to the administration order of measures were documented, although the specifics for each individual were too detailed to be presented in this dissertation. The varying delayed recall memory intervals that resulted from these adaptations were presented. However, variations due to changing distracter measures were only one factor affecting the very wide range of delay times that were observed.

Summary of Evaluation of Research Programme

Each of the 20 objectives outlined at the beginning of this research programme were met. This research was conducted in an ethical manner, which considered carefully the specific needs of participants, particularly those with severe disabilities, and did so in a manner that maintained scientific rigour and objectivity.

Brooks (1989) stated, “There are no untestable patients, only incompetent psychologists” (p. 82). However, this research programme identified that competent psychologists have few empirically-based guidelines for conducting assessments with severely disabled individuals. Within the limitations previously discussed, the research programme has made a significant contribution to the knowledge of this part of the field of neuropsychology, providing clinically relevant information about working with people with severe expressive disabilities.

REFERENCES

Accident Rehabilitation and Compensation Insurance Corporation. (1993). *ACC injury statistics 1992.* Wellington, New Zealand: ACC Statistical Publications.

Acimovic, M.L., Keatley, M.A., & Lemmon, J. (1993). The importance of qualitative indicators in the assessment of mild brain injury. *The Journal of Cognitive Rehabilitation, Nov/Dec,* 8-14.

Adams, K.M., & Heaton, R.K. (1990). Statement concerning the NIMH neuropsychological battery. *Journal of Clinical and Experimental Neuropsychology, 12*, 960-962.

American Psychological Association. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

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

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17*, 143-150.

Army Individual Test Battery. (1944). *Manual of directions and scoring.* Washington, D.C.: War Department, Adjutant General's Office.

Arnett, P.A., Rao, S.M., Bernardin, L., Grafman, J., Yetkin, F.Z., & Lobeck, L. (1994). Relationship between frontal lobe lesions and Wisconsin Card Sorting Test performance in patients with multiple sclerosis. *Neurology, 44,* 420-425.

Axelrod, B.N., Goldman, R.S., Heaton, R.K., Curtiss, G., Thompson, L.L., Chelune, G.J., & Kay, G.G. (1996). Discriminability of the Wisconsin Card Sorting Test using the standardization sample. *Journal of Clinical and Experimental Neuropsychology, 18,* 338-342.

Axelrod, B.N., & Millis, S.R. (1994). Preliminary standardization of the Cognitive Estimation Test. *Assessment, 1,* 269-274.

Babbage, D.R., & Leathem, J.M. (1997, September). *Neuropsychological assessment: Difficulties associated with "hard-to-assess" clients.* Paper presented at the New Zealand Psychological Society Conference, Palmerston North, New Zealand.

Babbage, D.R., & Leathem, J.M. (1998a). Neuropsychological assessment services for individuals with physical and sensory disabilities. *Traumatic Brain Injury: Effects and Outcomes. Proceedings of the Head Injury Society of New Zealand Conference 1998.* Wellington, New Zealand: Head Injury Society of New Zealand.

Babbage, D.R., & Leathem, J.M. (1998b, August). Neuropsychological assessment of "hard-to-assess" clients: Practitioner survey. Paper presented at the New Zealand Psychological Society / New Zealand College of Clinical Psychologists Conference, Wellington, New Zealand.

Babbage, D.R., & Leathem, J.M. (1999a, June). *Assessment practices of neuropsychologists for individuals with physical and sensory disabilities.* Poster session presented at the 3rd World Congress on Brain Injury, Quebec City, Quebec, Canada.

Babbage, D.R., & Leathem, J.M. (1999b, August). *Assessment of cognitive functioning in individuals with expressive disabilities in addition to traumatic brain injuries.* Paper presented at the 17th Annual Australasian Winter Conference on Brain Research, Queenstown, New Zealand.

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

Babcock, H., & Levy, L. (1940). *The measurement of efficiency of mental functioning (revised examination): Test and manual of directions.* Chicago: C.H. Stoelting.

Beaumont, J.G., Marjoribanks, J., Flury, S., & Lintern, T. (1999). Assessing auditory comprehension in the context of severe physical disability: the PACST. *Brain Injury, 13* (2), 99-112.

Beck, A.T. (1987). *Beck Depression Inventory.* San Antonio, TX: The Psychological Corporation.

Beck, A.T., Sterer, R.A., & Brown, G.K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Benton, A. (1992). Clinical neuropsychology: 1960-1990. *Journal of Clinical and Experimental Neuropsychology, 14,* 407-417.

Benton, A.L. (1959). *Right-Left Discrimination and Finger Localization.* New York: Hoeber.

Benton, A.L. (1974). *Revised Visual Retention Test: Clinical and experimental applications* (4th ed.). New York, NY: The Psychological Corporation.

Benton, A.L., & Hamsher, K. de S. (1989). *Multilingual Aphasia Examination.* Iowa City, IA: AJA Associates.

Benton, A.L., Hamsher, K. deS., Rey, G.J., & Sivan, A.B. (1994). *Multilingual Aphasia Examination* (3rd ed., pp. 165-213). Iowa City, IA: AJA Associates.

Benton, A.L., Hamsher, K. deS., Varney, N.R., & Spreen, O. (1983). *Contributions to neuropsychological assessment.* New York: Oxford University Press.

Benton, A.L., Hannay, H.J., & Varney, N.R. (1975). Visual perception of line direction in patients with unilateral brain disease. *Neurology, 25,* 907-910.

Benton, A.L., Sivan, A.B., Hamsher, K. de S., Varney, N.R., & Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual* (2nd ed.). New York: Oxford University Press.

Benton, A.L., & Tranel, D. (1993). Visuoperceptual, visuospatial, and visuoconstructive disorders. In K.M. Heilmann & E. Valenstein (Eds.), *Clinical neuropsychology* (3rd ed.). New York: Oxford University Press.

Benton, A.L., Van Allen, M.W., & Fogel, M.L. (1964). Temporal orientation in cerebral disease. *Journal of Nervous and Mental Disease, 139,* 110-119.

Berent, S., Giordani, B., Gilman, S., Junck, L., Lehtinen, S., Kluin, K., & Koeppe, R. (1992). Methodologic considerations in neuropsychological testing of ataxic patients: In reply. *Archives of Neurology, 49,* 218-219.

Berent, S., Giordani, B., Gilman, S., Junck, L., Lehtinen, S., Markel, D.S., Boivin, M., Kluin, K.J., Parks, P., & Koeppe, R.A. (1990). Neuropsychological changes in olivopontocerebellar atrophy. *Archives of Neurology, 47*, 997-1001.

Berninger, V.W., Gans, B.M., St. James, P., & Connors, T. (1988). Modified WAIS-R for patients with speech and/or hand dysfunction. *Archives of Physical Medicine and Rehabilitation, 69,* 250-255.

Bigge, J. (1982). Severe communication problems. In J.L. Bigge (Ed.), *Teaching individuals with physical and multiple disabilities* (2nd ed., pp. 75-109). Columbus, OH: Charles E. Merril Publishing Company.

Bigler, E.D., Rosa, L., Schultz, F., Hall, S., & Harris, J. (1989). Rey-Auditory Verbal Learning and Rey-Osterreith Complex Figure Design performance in Alzheimer's disease and closed head injury. *Journal of Clinical Psychology, 45,* 277-280.

Bigler, E.D., & Tucker, D.M. (1981). Comparison of Verbal IQ, Tactual Performance, Seashore Rhythm and Finger Oscillation tests in the blind and brain-damaged. *Journal of Clinical Psychology, 37*, 849-851.

Binder, L.M. (1997). A review of mild head trauma. Part II: Clinical implications. *Journal of Clinical and Experimental Neuropsychology, 19*, 432-457.

Binder, L.M., Rohling, M.L., & Larrabee, G.J. (1997). A review of mild head trauma. Part I: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology, 19*, 421-431.

Binder, L.M., & Thompson, L.L. (1995). The ethics code and neuropsychological assessment practices. *Archives of Clinical Neuropsychology, 10*, 27-46.

Black, F.W. (1986). Neuroanatomic and neuropsychologic correlates of digit span performance by brain-damaged adults. *Perceptual and Motor Skills, 63,* 815-822.

Black, F.W., & Strub, R.L. (1978). Digit repetition performance in patients with focal brain damage. *Cortex, 14* (1), 12-21.

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

Blum, L.H., Burgemeister, B., & Lorge, I. (1951). Trends in estimating the mental maturity of the cerebral palsied child. *Journal of Exceptional Children, 17,* 174-177.

Boller, F., & Vignolo, L.A. (1966). Latent sensory aphasia in hemisphere-damaged patients: An experimental study with the Token Test. *Brain, 89,* 815-831.

Bond, J.A., & Buchtel, H.A. (1984). Comparison of the Wisconsin Card Sorting Test and the Halstead Category Test. *Journal of Clinical Psychology, 40,* 1251-1254.

Boring, E.G., (1923). Intelligence as the tests test it. *The New Republic,* June 6, pp. 35-37.

Botterbusch, K.F. (1976). *The use of psychological tests with individuals who are severely disabled.* Menomonee, WN: Materials Development Center, Stout Vocational Rehabilitation Institute.

Bowden, S.C., & Smith, L.C. (1994). What does the Austin maze measure? *Australian Psychologist, 29* (1), 34-37.

Boyd, J.L. (1981). A validity study of the Hooper Visual Organization Test. *Journal of Consulting and Clinical Psychology, 49*, 15-19.

Boyd, J.L. (1982a). Reply to Rathburn and Smith: Who made the Hooper blooper? *Journal of Consulting and Clinical Psychology, 50*, 284-285.

Boyd, J.L. (1982b). Reply to Woodward. *Journal of Consulting and Clinical Psychology, 50*, 289-290.

Brandon, A.D., Chavez, E.L., & Bennett, T.L. (1986). A comparative evaluation of two neuropsychological finger tapping instruments: Halstead-Reitan and Western Psychological Services. *The International Journal of Clinical Neuropsychology, 8* (2), 64-65.

Brooks, N. (1989). Closed head trauma: Assessing the common cognitive problems. In M.D. Lezak (Ed.), *Assessment of the behavioral consequences of head trauma* (pp. 61-85). New York: Alan R. Liss, Inc.

Buschke, H., & Fuld, P.A. (1974). Evaluation of storage, retention, and retrieval in disordered memory and learning. *Neurology, 11,* 1019-1025.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory (MMPI-2). Manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Butler, M., Retzlaff, P., & Vanderploeg, R. (1991). Neuropsychological test usage. *Professional Psychology: Research and Practice, 22*, 510-512.

Butters, N., Grant, I., Haxby, J., Judd, L.L., Martin, A., McClelland, J., Pequegnat, W., Schacter, D., & Stover, E. (1990). Assessment of AIDS-related cognitive changes: Recommendations of the NIMH workshop on neuropsychological assessment approaches. *Journal of Clinical and Experimental Neuropsychology, 12,* 963-978.

Caplan, B., & Caffrey, D. (1992). Fractionating block design: Development of a test of visuospatial analysis. *Neuropsychology, 6,* 385-394.

Caplan, B., & Shechter, J. (1995). The role of nonstandard neuropsychological assessment in rehabilitation: History, rationale, and examples. In L.A. Cushman & M.J. Scherer (Eds.), *Psychological assessment in medical rehabilitation* (pp. 359-391). Washington, DC: American Psychological Association.

Carlesimo, G.A., & Caltagirone, C. (1995). Components in the visual processing of known and unknown faces. *Journal of Clinical and Experimental Neuropsychology, 17*, 691-705.

Cattell, R.B., & Cattell, A.K.S. (1960). *Culture Fair Intelligence Test.* Champaign, IL: Institute for Personality and Ability Testing.

Chelune, G.J., & Bornstein, R.A. (1988). WMS-R patterns among patients with unilateral brain lesions. *The Clinical Neuropsychologist, 2,* 121-132.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Clifton, G.L., Hayes, R.L., Levin, H.S., Michel, M.E., & Choi, S.C. (1992). Outcome measures for clinical trials involving traumatically brain-injured clients: Report of a conference. *Neurosurgery, 31*, 975-978.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hilsdale, NJ: Earlbaum Associates.

Coleman, A.R., Moberg, P.J., Ragland, J.D., & Gur, R.C. (1997). Comparison of the Halstead-Reitan and Infrared Light Bean finger tappers. *Assessment, 4*, 277-286.

Connolly, J.F., Mate-Kole, C.C., & Joyce, B.M. (1999). Global aphasia: An innovative assessment approach. *Archives of Physical Medicine and Rehabilitation, 80,* 1309-1315.

Coolican, H. (1994). *Research methods and statistics in psychology* (2nd ed.). London: Hodder & Stoughton.

Coop, R.H., Eckel, E., & Stuck, G.B. (1975). An assessment of the pictorial test of intelligence for use with young cerebral-palsied children. *Developmental Medicine and Child Neurology, 17,* 287-292.

Correll, R.E., Brodginski, S.E., & Rokosz, S.F. (1993). WAIS performance during the acute recovery stage following closed-head injury. *Perceptual and Motor Skills, 76,* 99-109.

Crossen, J.R., & Wiens, A.N. (1994). Comparison of the Auditory–Verbal Learning Test (AVLT) and California Verbal Learning Test (CVLT) in a sample of normal subjects. *Journal of Clinical and Experimental Neuropsychology, 16*, 190-194.

Culver, C.M. (1969). Test of right-left discrimination. *Perceptual and Motor Skills, 29,* 863-867.

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

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

Denman, S.B. (1984). *Denman Neuropsychology Memory Scale.* Charleston, SC: S.B. Denman.

Denman, S.B. (1987). *Denman Neuropsychology Memory Scale: Norms.* Charleston, SC: Sidney B. Denman.

Derogatis, L.R. (1994). *Symptom Checklist-90-R: Administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems, Inc.

Dikmen, S.S., Machamer, J.E., Winn, H.R., & Temkin, N.R. (1995). Neuropsychological outcome at 1-yaer post head injury. *Neuropsychology, 9*, 80-90.

Dodrill, C.B. (1997). Myths of neuropsychology. *The Clinical Neuropsychologist, 11*, 1-17.

Dorman, C., Hurley, A.D., & Laatsch, L. (1984). Prediction of spelling and reading performance in cerebral palsied adolescents using neuropsychological tests. *The International Journal of Clinical Neuropsychology, 6*, 142-144.

Dorman, C., Laatsch, L.K., & Hurley, A.D. (1985). The applicability of neuropsychological test batteries for assessment of the congenitally brain disordered. *The International Journal of Clinical Neuropsychology, 7*, 111-117.

Dunn, L.M., & Dunn, E.S. (1981). *Peabody Picture Vocabulary Test-Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L.M., & Markwardt, F.C. (1970). *Peabody Individual Achievement Test Manual.* Circle Pines, MN: American Guidance Service.

Ebner, D.J., Burkhead, E.J., & Merydith, S.P. (1994). The Americans with Disabilities Act: Implications for vocational assessment. *Assessment in Rehabilitation and Exceptionality, 1* (2), 91-97.

Elwood, R.W. (1993). Clinical discriminations and neuropsychological tests: An appeal to Bayes' theorem. *The Clinical Neuropsychologist, 7*, 224-233.

Embretson, S.E. (1996). The new rules of measurement. *Psychological Assessment, 8*, 341-349.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.

Faust, D. (1991). Forensic neuropsychology: The art of practicing a science that does not yet exist. *Neuropsychology Review, 2*, 205-229.

Faust, D., Guilmette, T.J., Hart, K., Arkes, H.R., Fishburne, F.J., & Davey, L. (1988). Neuropsychologists' training, experience, and judgment accuracy. *Archives of Clinical Neuropsychology, 3,* 145-163.

Finlayson, M.A.J., & Reitan, R.M. (1980). Effect of lateralized lesions on ipsilateral and contralateral motor functioning. *Journal of Clinical Neuropsychology, 2,* 237-243.

Fitz, A.G., Conrad, P.M., Hom, D.L., Sarff, P., & Marjovski, L.V. (1992). Hooper Visual Organization Test performance in lateralized brain injury. *Archives of Clinical Neuropsychology, 7,* 243-250.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state." *Journal of Psychiatric Research, 12,* 189-198.

Francis, D.J., Fletcher, J.M., Rourke, B.P., & York, M.J. (1992). A five-factor model for motor, psychomotor, and visual-spatial tests used in the neuropsychological assessment of children. *Journal of Clinical and Experimental Neuropsychology, 14*, 625-637.

Franzen, M.D. (with Robbins, D.E., & Sawicki, R.F.). (1989). *Reliability and validity in neuropsychological assessment.* New York: Plenum Press.

Gabrys, J.B. (1979). The Babcock Story Recall of behaviorally disordered children scoring high or low on extraversion. *Perceptual and Motor Skills, 48,* 157-158.

Garb, H.N., & Schramke, C.J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analyses. *Psychological Bulletin, 120*, 140-153.

Gardner, M.F. (1982). *TVPT: Test of Visual-Perceptual Skills (non-motor).* Los Angeles, CA: Western Psychological Services.

Gass, C.S., & Brown, M.C. (1992). Neuropsychological test feedback to patients with brain dysfunction. *Psychological Assessment, 4*, 272-277.

Geffen, G.M., Butterworth, P., & Geffen, L.B. (1994). Test-retest reliability of a new form of the Auditory Verbal Learning Test (AVLT). *Archives of Clinical Neuropsychology, 9*, 303-316.

Geisinger, K.F., & Carlson, J.F. (1995). Standards and standardization. In J.N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 211-223). New York: Oxford University Press.

Gerson, A. (1974). Validity and reliability of the Hooper Visual Organization Test. *Perceptual and Motor Skills, 39*, 95-100.

Golden, C.J. (1978). *Stroop Color and Word Test.* Chicago, IL: Stoelting.

Golden, C.J., & Robbins, D.E. (1985). Considerations in cases of visual, auditory, or motor impairment. In D.P. Swiercinsky (Ed.), *Testing adults: A reference guide for special psychodiagnostic assessments* (pp. 89-100). Kansas City: Test Corporation of America.

Goldstein, G. (1987). Neuropsychological assessment for rehabilitation: Fixed batteries, automated systems, and non-psychometric methods. In M.J. Meier, A.L. Benton, & L. Diller (Eds.), *Neuropsychological rehabilitation* (pp. 18-40). Edinburgh: Churchill Livingstone.

Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination.* Philadelphia: Lea & Febiger.

Gorham, D.R. (1956a). *Clinical manual for the Proverbs Test.* Missoula, MT: Psychological Test Specialists.

Gorham, D.R. (1956b). A Proverbs Test for clinical and experimental use. *Psychological Reports, 2,* 1-12.

Grant, D.A., & Berg, E.A. (1948). A behavioral analysis of the degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology, 55,* 404-411.

Grimm, L.G., & Yarnold, P.R. (Eds.). (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Gronwall, D. (1977). Paced Auditory Serial Addition Task: A measure of recovery from concussion. *Perceptual and Motor Skills, 44,* 367-373.

Gronwall, D., & Sampson, H. (1974). *The psychological effects of concussion.* Auckland, New Zealand: University Press/Oxford University Press.

Gronwall, D.M.A. (1977). Paced Auditory Serial-Addition Task: A measure of recovery from concussion. *Perceptual and Motor Skills, 44,* 367-373.

Guilmette, T.J., Faust, D., Hart, K., & Arkes, H.R. (1990). A national survey of psychologists who offer neuropsychological services. *Archives of Clinical Neuropsychology, 5,* 373-392.

Guilmette, T.J., & Giuliano, A.J. (1991). Taking the stand: Issues and strategies in forensic neuropsychology. *The Clinical Neuropsychologist, 5*, 197-219.

Haaland, K.Y., Temkin. N., Randahl, G., & Dikmen, S. (1994). Recovery of simple motor skills after head injury. *Journal of Clinical and Experimental Neuropsychology, 16,* 448-456.

Hair, J.F., Jr., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall International.

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

Hamilton, M.A. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56-62.

Hamilton, M.A. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6,* 278-296.

Hamsher, K. de S., & Roberts, R.J. (1985). Memory for recent U.S. presidents in patients with cerebral disease. *Journal of Clinical and Experimental Neuropsychology, 7,* 1-13.

Hannay, H.J., & Levin, H.S. (no date). *Continuous Recognition Memory Test.* Available from H.J. Hannay, 4046 Grenock, Houston, TX 77025.

Hannay, H.J., Levin, H.S., & Grossman, R.G. (1979). Impaired recognition memory after head injury. *Cortex, 15,* 269-283.

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded* (Rev. ed.). Odessa, FL: Psychological Assessment Resources.

Heaton, R.K., Grant, I., & Matthews, C.G. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery: Demographic corrections, research findings, and clinical applications.* Odessa, FL: Psychological Assessment Resources.

Heaton, S.K., & Heaton, R.K. (1981). Testing the impaired patient. In S.B. Filskov & T.J. Boll (Eds.), *Handbook of clinical neuropsychology* (pp. 526-544). New York: John Wiley and Sons.

Heilbronner, R.L. (1992). The search for a "pure" visual memory test: Pursuit of perfection? *The Clinical Neuropsychologist, 6*, 105-112.

Heilbronner, R.L., Buck, P., & Adams, R.L. (1989). Factor analysis of verbal and nonverbal clinical memory tests. *Archives of Clinical Neuropsychology, 4,* 299-309.

Heinrichs, R.W. (1990). Current and emergent applications of neuropsychological assessment: Problems of validity and utility. *Professional Psychology: Research and Practice, 21*, 171-176.

Hinton-Bayre, A.D., Geffen, G., & McFarland, K. (1996). Impairment and recovery of speed of information processing after mild head injury: A prospective study of rugby league players. In J. Ponsford, P. Snow, & V. Anderson (Eds.), *International perspectives in traumatic brain injury: Proceedings of the 5th conference of the International Association for the Study of Traumatic Brain Injury and 20th conference of the Australian Society for the Study of Brain Impairment* (pp. 291-298). Bowen Hills, Australia: Australian Academic Press.

Hom, J., & Reitan, R.M. (1990). Generalized cognitive function after stroke. *Journal of Clinical and Experimental Neuropsychology, 12,* 644-654.

Hooper, H.E. (1958). *The Hooper Visual Organization Test manual.* Los Angeles: Western Psychological Services.

Hooper, H.E. (1983). *Hooper Visual Organization Test (VOT).* Los Angeles: Western Psychological Services.

Ivnik, R.J., Malec, J.F., Smith, G.E., Tangalos, E.G., & Petersen, R.C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist, 10*, 262-278.

Ivnik, R.J., Smith, G.E., Lucas, J.A., Tangalos, E.G., Kokmen, E., & Petersen, R.C. (1997). Free and Cued Recall Selective Reminding Test: MOANS norms. *Journal of Clinical and Experimental Neuropsychology, 19*, 676-691.

Johnstone, B., & Wilhelm, K.L. (1997). The construct validity of the Hooper Visual Organization Test. *Assessment, 4*, 243-248.

Jones, W.P. (1989). A proposal for the use of Bayesian probabilities in neuropsychological assessment. *Neuropsychology, 3,* 17-22.

Jordan, S. (1978). Haptic intelligence scale for adult blind. *Perceptual and Motor Skills, 47*, 203-222.

Kaplan, E., Fein, D., Morris, R., & Delis, D.C. (1991). *WAIS-R as a neuropsychological instrument: WAIS-R-NI manual.* New York: The Psychological Corporation.

Kaplan, E.F., Goodglass, H., & Weintraub, S. (1978). *The Boston Naming Test* (Experimental ed.). Boston: Kaplan & Goodglass.

Kaplan, E.F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Philadelphia: Lea & Febiger.

Kaszniak, A.W. (1986). Neuropsychology of dementia. In I. Grant & K. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 172-220). New York: Oxford University Press.

Kempler, D., & Van Lancker, D. (1987). The right turn of phrase. *Psychology Today, 21* (4), 20-22.

Kempler, D., & Van Lancker, D. (1996). *Familiar and Novel Language Comprehension Test.* Los Angeles, CA: Author.

Kempler, D., Van Lancker, D., Marchman, V., & Bates, E. (1999). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology, 15*, 327-349.

Kempler, D., Van Lancker, D., & Read, S. (1988). Proverb and idioms interpretation in Alzheimer Disease. *Alzheimer Disease and Associated Disorders, 2,* 38-49.

Kerlinger, F. (1973). *Foundations of behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.

Kiernan, R.J. (1989). The problem for neuropsychologists: Differentiating physical, emotional, and cognitive abilities. In B.W. Heller, L.M. Flohr, & L.S. Zegans (Eds.), *Psychosocial interventions with physically disabled persons* (pp. 58-77). New Brunswick, NJ: Rutgers University Press.

Kirk, S.A., McCarthy, J.J., & Kirk, W.D. (1968). *Illinois Test of Psycholinguistic Abilities, Revised.* Wood Dale, IL: Stoelting.

Kleinmuntz, B. (1990). Why we still use our heads instead of the formulas: Toward an integrative approach. *Psychological Bulletin, 107,* 296-310.

Kopelman, M.D., Wilson, B.A., & Baddeley, A.D. (1989). The Autobiographical Memory Interview: A new assessment of autobiographical and personal semantic memory in amnesic patients. *Journal of Clinical and Experimental Neuropsychology, 11,* 724-744.

Kreutzer, J.S., Bale, P., Chase, J., Lourie, I., Morrison, K., & Schools, L. (1986, August). *The Babcock Story Recall Test: Inter-rater reliability and normative data.* Paper presented at the meeting of the American Psychological Association, Washington, DC.

Kreutzer, J.S., Gordon, W.A., Rosenthal, M., & Marwitz, J. (1993). Neuropsychological characteristics of patients with brain injury: Preliminary findings from a multicenter investigation. *Journal of Head Trauma Rehabilitation, 8* (2), 47-59.

Kupke, T., Lewis, R., & Rennick, P.M. (1979). Sex differences in the neuropsychological functioning of epileptics. *Journal of Consulting and Clinical Psychology, 47*, 1128-1130.

Laine, M., & Butters, N. (1982). A preliminary study of the problem-solving strategies of detoxified long-term alcoholics. *Drug and Alcohol Dependence, 10,* 235-242.

Lalonde, R., Botez, T., & Botez, M.I. (1992). Methodologic considerations in neuropsychological testing of ataxic patients. *Archives of Neurology, 49*, 218.

Larrabee, G.J., Largen, J.W., & Levin, H.S. (1985). Sensitivity of age-decline resistant ("Hold") WAIS subtests to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 7,* 497-504.

Leininger, B.E., Gramling, S.E., Farrell, A.D., Kreutzer, J.S., & Peck, E. (1990). Neuropsychological deficits in symptomatic minor head injury patients after concussion and mild concussion. *Journal of Neurology, Neurosurgery, and Psychiatry, 53,* 293-296.

Leonberer, F.T., Nicks, S.D., Goldfader, P.R., & Munz, D.C. (1991). Factor analysis of the Wechsler Memory Scale–Revised and the Halstead-Reitan Neuropsychological Battery. *The Clinical Neuropsychologist, 5,* 83-88.

Leonberger, F.T., Nicks, S.D., Larrabee, G.J., & Goldfader, P.R. (1992). Factor structure of the Wechsler Memory Scale–Revised within a comprehensive neuropsychological battery. *Neuropsychology, 6,* 239-249.

Levin, B.E., Llabre, M.M., & Reisman, S. (1991). Visuospatial impairment in Parkinson's disease. *Neurology, 41,* 365-369.

Levin, H.S. (1994). A guide to clinical neuropsychological testing. *Archives of Neurology, 51,* 854-859.

Levin, H.S., Gary, H.E., Eisenberg, H.M., Ruff, R.M., Barth, J.T., Kreutzer, J., High, W.M., Jr., Portman, S., Foulkes, M.A., Jane, J.A., Marmarou, A., & Marshall, L.F. (1990). Neurobehavioral outcome 1 year after severe head injury: Experience of the Traumatic Coma Data Bank. *Journal of Neurosurgery, 73,* 699-709.

Levin, H.S., O'Donnell, V.M., & Grossman, R.G. (1979). The Galveston Orientation and Amnesia Test: A practical scale to assess cognition after head injury. *Journal of Nervous and Mental Disease, 167,* 675-684.

Levin, H.S., Song, J., Scheibel, R.S., Fletcher, J.M., Harward, H., Lilly, M., & Goldstein, F. (1997). Concept formation and problem-solving following closed head injury in children. *Journal of the International Neuropsychological Society, 3,* 598-607.

Lewis, R., & Kupke, T. (1992). Intermanual differences on skilled and unskilled motor tasks in nonlateralized brain dysfunction. *The Clinical Neuropsychologist, 6,* 374-382.

Lewis, S., Campbell, A., Takushi-Chinen, R., Brown, A., Dennis, G., Wood, D., & Weir, R. (1997). Visual Organization Test performance in an African American population with acute unilateral cerebral lesions. *International Journal of Neuroscience, 91*, 295-302.

Lezak, M.D. (1978). Subtle sequelae of brain damage: Perplexity, distractibility, and fatigue. *American Journal of Physical Medicine, 57* (1), 9-15.

Lezak, M.D. (1988). IQ: R.I.P. *Journal of Clinical and Experimental Neuropsychology, 10*, 351-361.

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

Libon, D.J., Mattson, R.E., Glosser, G., Kaplan, E., Malamut, M., Sands, L.P., Swenson, R., & Cloud, B.S. (1996). A nine-word dementia version of the California Verbal Learning Test. *Clinical Neuropsychologist, 10,* 237-244.

Likert, R., & Quasha, W.H. (1970). *The revised Minnesota Paper Form Board Test.* New York: The Psychological Corporation.

MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.

Magner, J.R., & Spector, J. (1984). Adaptation of neuropsychological assessment procedures for a profoundly response-limited patient. *Cognitive Rehabilitation, 2* (5), 28-31.

Mapou, R.L. (1988). Testing to detect brain damage: An alternative to what may no longer be useful. *Journal of Clinical and Experimental Neuropsychology, 10*, 271-278.

Margolin, D.I., Pate, D.S., Friedrich, F.J., & Elia, E. (1990). Dysnomia in dementia and in stroke patients: Different underlying cognitive deficits. *Journal of Clinical and Experimental Neuropsychology, 12,* 597-612.

Markwardt, F.C. (1989). *Peabody Individual Achievement Test-Revised.* Circle Pines, MN: American Guidance Service.

Matthews, C.G., & Kløve, H. (1964). *Instruction manual for the Adult Neuropsychology Test Battery.* Madison, WI: University of Wisconsin Medical School.

Mayberry, R.I. (1992). The cognitive development of deaf children: Recent insights. In S.J. Segalowitz & I. Rapin (Eds.), *Handbook of neuropsychology, Vol. 7: Child neuropsychology* (pp. 51-68). Amsterdam: Elsevier.

McCaffrey, R.J., Ortega, A., & Haase, R.F. (1993). Effects of repeated neuropsychological assessments. *Archives of Clinical Neuropsychology, 8,* 519-524.

McCaffrey, R.J., Ortega, A., Orsillo, S.M., Nelles, W.B., & Haase, R.F. (1992). Practice effects in repeated neuropsychological assessments. *The Clinical Neuropsychologist, 6*, 32-42.

McCaffrey, R.J., & Westervelt, H.J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review, 5*, 203-221.

McMillan, T.M. (1997). Neuropsychological assessment after extremely severe head injury in a case of life or death. *Brain Injury, 11*, 483-490.

Microsoft Corporation. (1998). *Microsoft Excel 98 for Macintosh.* Redmond, WA: Author.

Millis, S.R., Malina, A.C., Bowers, D.A., & Ricker, J.H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale-III. *Journal of Clinical and Experimental Neuropsychology, 21*, 87-93.

Milner, B. (1965). Visually-guided maze learning in man: Effects of bilateral hippocampal, bilateral frontal, and unilateral cerebral lesions. *Neuropsychologia, 3,* 317-338.

Minden, S.L., Moes, E.J., Orav, J., Kaplan, E., & Reich, P. (1990). Memory impairment in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology, 12,* 566-586.

Mitrushina, M., & Satz, P. (1995). Repeated testing of normal elderly with the Boston Naming Test. *Aging: Clinical and Experimental Research, 7,* 123-127.

Mitrushina, M.N., Boone, K.B., & D'Elia, L.F. (1999). *Handbook of normative data for neuropsychological assessment*. New York: Oxford University Press.

Mittler, P. (1970). *The psychological assessment of mental and physical handicaps.* London: Tavistock Publications/Methuen & Co Ltd.

Moehle, K.A., Rasmussen, J.L., & Fitzhugh-Bell, K.B. (1990). Factor analysis of neuropsychological tests in an adult sample. *International Journal of Clinical Neuropsychology, 12*, 107-115.

Montgomery, G.K. (1995). A multi-factor account of disability after brain injury: implications for neuropsychological counseling. *Brain Injury, 9*, 453-469.

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

Mountain, M.A., & Snow, W.G. (1993). Wisconsin Card Sorting Test as a measure of frontal pathology: A review. *The Clinical Neuropsychologist, 7,* 108-118.

Naugle, R.I. (1990). Epidemiology of traumatic brain injury in adults. In E.D. Bigler (Ed.), *Traumatic brain injury: Mechanisms of damage, assessment, intervention and outcome* (pp. 69-103). Austin, TX: Pro-Ed.

Nelson, H.E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313-324.

Nelson, H.E. (1982). *The National Adult Reading Test (NART): Test manual.* Windsor, United Kingdom: NFER-Nelson.

Nelson, L.D., Satz, P., Mitrushina, M., van Gorp, W., Cicchetti, D., Lewis, R., & van Lancker, D. (1989). Development and validation of the Neuropsychology Behavior and Affect Profile. *Psychological Assessment, 1,* 266-272.

O'Donnell, J.P., Radtke, R.C., Leicht, D.J., & Caesar, R. (1988). Encoding and retrieval processes in learning-disabled, head-injured, and nondisabled young adults. *The Journal of General Psychology, 115,* 355-368.

Osterreith, P.A. (1993). The complex figure copy test. *The Clinical Neuropsychologist, 7*, 3-21 (J. Corwin & F.W. Bylsma, Trans.). (Original work published 1944)

Palmer, B.W., Heaton, R.K., Paulsen, J.S., Kuck, J., Braff, D., Harris, M.J., Zisook, S., & Jeste, D.V. (1997). Is it possible to be schizophrenic yet neuropsychologically normal? *Neuropsychology, 11*, 437-446.

Paolo, A.M., Cluff, B.R., & Ryan, J.J. (1996). Influence of perceptual organization and naming abilities on the Hooper Visual Organization Test. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 9*, 254-257.

Paraskevopoulos, J.N., & Kirk, S.A. (1969). *The development and psychometric characteristics of the Revised Illinois Test of Psycholinguistic Abilities.* Urbana, IL: University of Illinois Press.

Parsons, O.A. (1991). Clinical neuropsychology 1970-1990: A personal view. *Archives of Clinical Neuropsychology, 6,* 105-111.

Peach, R.K. (1992). Factors underlying neuropsychological test performance in chronic severe traumatic brain injury. *Journal of Speech and Hearing Research, 35,* 810-818.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologica, 20,* 249-262.

Prigatano, G.P., Fordyce, D.J., Zeiner, H.K., Roueche, I.R., Pepping, M., & Wood, B.C. (1986). *Neuropsychological rehabilitation after brain injury.* Baltimore: Johns Hopkins University Press.

Prigatano, G.P., & Redner, J.E. (1993). Uses and abuses of neuropsychological testing in behavioral neurology. *Behavioral Neurology, 11*, 219-231.

Puente, A.E. (1990). Psychological assessment of minority group members. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (2nd ed., pp. 505-520). New York: Pergamon.

Puente, A.E. (1992). The status of clinical neuropsychology. *Archives of Clinical Neuropsychology, 7,* 297-312.

Putnam, S.H., & DeLuca, J.W. (1990). The TCN professional practice survey: Part I: General practices of neuropsychologists in primary employment and private practice settings. *The Clinical Neuropsychologist, 4*, 199-244.

Randt, C.T., & Brown, E.R. (1986). *Randt Memory Test.* Bayport, NY: Life Science Associates.

Rapaport, D., Gill, M.M., & Schafer, R. (1968). *Diagnostic psychological testing* (Rev. ed.) (R.R. Holt, Ed.). New York: International University Press.

Rathburn, J., & Smith, A. (1982). Comment on the validity of Boyd's validation study of the Hooper Visual Organization Test. *Journal of Consulting and Clinical Psychology, 50,* 281-283.

Raven, J.C. (1960). *Guide to the Standard Progressive Matrices.* London: H.K. Lewis.

Raven, J.C. (1965). *Guide to using the Coloured Progressive Matrices.* London: H.K. Lewis.

Raven, J.C., Court, J.H., & Raven, J. (1976). *Manual for Raven's Progressive Matrices.* London: H.K. Lewis.

Reitan, R.M., & Davison, L.A. (1974). *Clinical neuropsychology: Current status and applications.* Washington, DC: Winston.

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

Reitan, R.M., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation.* Tuscon, AZ: Neuropsychology Press.

Reitan, R.M., & Wolfson, D. (1994a). Dissociation of motor impairment and higher-level brain deficits in strokes and cerebral neoplasms. *The Clinical Neuropsychologist, 8,* 193-208.

Reitan, R.M., & Wolfson, D. (1994b). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4*, 161-198.

Reitan, R.M., & Wolfson, D. (1995). Influence of age and education on neuropsychological test results. *The Clinical Neuropsychologist, 9*, 151-158.

Reitan, R.M., & Wolfson, D. (1996). Relationships between specific and general tests of cerebral functioning. *The Clinical Neuropsychologist, 10,* 37-42.

Retzlaff, P.D., & Gibertini, M. (1994). Neuropsychometric issues and problems. In R.D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (pp. 185-209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rey, A.R. (1964). *L'examen clinique en psychologie.* Paris: Presses Universitaires de France.

Rey, A.R. (1993). Psychological examination of traumatic encephalopathy. *The Clinical Neuropsychologist, 7*, 3-21 (J. Corwin & F.W. Bylsma, Trans.). (Original work published 1941)

Rezai, K., Andeasen, N.C., Alliger, R., Cohen, G., Swayze, V., II, & O'Leary, D.S. (1993). The neuropsychology of the prefrontal cortex. *Archives of Neurology, 59,* 636-642.

Richards, J.S., Brown, L., Hagglund, K., Bua, G., & Reeder, K. (1988). Spinal cord injury and concomitant traumatic brain injury. *American Journal of Physical Medicine and Rehabilitation, 67,* 211-216.

Ricker, J.H., & Axelrod, B.N. (1995). Hooper Visual Organization Test: Effects of object naming ability. *The Clinical Neuropsychologist, 9*, 57-62.

Robinette, R.L., Sherer, M., & Adams, R.L. (1993). The development of verbal and nonverbal factorially derived memory measures. *Journal of Clinical Psychology, 49*, 89-94.

Roth, E., Davidoff, G., Thomas, P., Doljanac, R., Dijkers, M., Berent, S., Morris, J., Yarkony, G. (1989). A controlled study of neuropsychological deficits in acute spinal cord injury patients. *Paraplegia, 27,* 480-489.

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

Russell, E.W., & Starkey, R. (1993). *Halstead-Russell Neuropsychological Evaluation System (HRNES).* Los Angeles, CA: Western Psychological Services.

Rust, J., & Golombok, S. (1989). *Modern psychometrics: The science of psychological assessment.* London: Routledge.

Sattler, J.M., & Tozier, L.L. (1970). A review of intelligence test modifications used with cerebral palsied and other handicapped groups. *The Journal of Special Education, 4,* 391-398.

Sbordone, R.J. (1992). Neuropsychological assessment of the traumatically brain injured client. *Psychotherapy in Private Practice, 10* (4), 45-52.

Sbordone, R.J., & Long, C.J. (Eds.). (1996). *Ecological validity of neuropsychological testing.* Delray Beach, FL: GR Press/St. Lucie Press.

Schmidt, F.L. (1996). Statistical significant testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115-129.

Schmidt, F.L., & Hunter, J.E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*, 199-223.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350-353.

Schonell, F.E. (1958). Intelligence testing. In R.S. Illingsworth (Ed.), *Recent advances in cerebral palsy* (pp. 132-145). London: J. & A. Churchill, Ltd.

Schretlen, D., Bobholz, J.H., & Brandt, J. (1996). Development and psychometric properties of the Brief Test of Attention. *The Clinical Neuropsychologist, 10,* 80-89.

Schretlen, D., Brandt, J., & Bobholz, J.H. (1996). Validation of the Brief Test of Attention in patients with Huntington's Disease and amnesia. *The Clinical Neuropsychologist, 10,* 90-95.

Schuler, A.L., & Perez, L. (1991). Assessment: Current concerns and future directions. In L.H. Meyer, C.A. Peck, & L. Brown (Eds.), *Critical issues in the lives of people with severe disabilities* (pp. 101-106). Baltimore, MA: Paul H. Brooks Publishing Company.

Schweinberger, S.R., Buse, C., Freeman, R.B., Jr., Schönle, P.W., & Sommer, W. (1992). Memory search for faces and digits in patients with unilateral brain lesions. *Journal of Clinical and Experimental Neuropsychology, 14*, 839-856.

Seashore, C.E., Lewis, D., & Saetveit, D.L. (1960). *Seashore Measures of Musical Talents* (Rev. ed.). New York: The Psychological Corporation.

Segall, M.H. (1986). Culture and behavior: Psychology in global perspective. *Annual Review of Psychology, 37,* 523-564.

Sellers, A.H., & Nadler, J.D. (1992). A survey of current neuropsychological assessment procedures used for different age groups. *Psychotherapy in Private Practice, 11* (3), 47-57.

Shallice, T., & Evans, M.E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 14,* 292-303.

Shepherd, I., & Leathem, J. (1999). Factors affecting performance in cross-cultural neuropsychology: From a New Zealand bicultural perspective. *Journal of the International Neuropsychological Society, 5,* 83-84.

Shiel, A., & Wilson, B.A. (1998). Assessment after extremely severe head injury in a case of life or death: Further support for McMillan. *Brain Injury, 12*, 809-816.

Shimoyama, I., Ninchoji, T., & Uemura, K. (1990). The Finger-Tapping Test: A quantitative analysis. *Archives of Neurology, 47,* 681-684.

Sinnett, E.R., & Holen, M.C. (1995). Chance performance in the absence of norms for neuropsychological tests: A clinical note. *Journal of Clinical Psychology, 51*, 400-402.

Sivan, A.B. (1992). *Benton Visual Retention Test* (5th ed.). San Antonio, TX: The Psychological Corporation.

Slick, D., Hopp, G., Strauss, E., Fox, D., Pinch, D., & Strickgold, K. (1996). Effects of prior testing with the WAIS-R NI on subsequent retest with the WAIS-R. *Archives of Clinical Neuropsychology, 11*, 123-130.

Smith, A. (1982). *Symbol Digit Modalities Test (SMDT) manual* (Rev. ed.). Los Angeles: Western Psychological Services.

Smith, S.T. Jr., Cash, C., Barr, S.E., & Putney, R.T. (1986). The nonspeech assessment of hemispheric specialization in retarded children. *Neuropsychologia, 24*, 293-296.

Sohlberg, M.M., & Mateer, C.A. (1989). *Introduction to cognitive rehabilitation.* New York: Guilford Press.

Speilberger, C.D. (1983). *State-Trait Anxiety Inventory (Form Y).* Palo Alto, CA: Mind Garden.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms and commentary* (2nd ed.). New York: Oxford University Press.

SPSS Inc. (1995). *SPSS for Macintosh 6.1.1.* Chicago, IL: Author.

Storandt, M., Botwinick, J., & Danziger, W.L. (1986). Longitudinal changes: Patients with mild SADT and matched healthy controls. In L.W. Poon (Ed.), *Handbook for clinical memory assessment of older adults* (pp. 277-284). Washington, DC: American Psychological Association.

Storandt, M., & Hill, R.D. (1989). Very mild senile dementia of the Alzheimer type: II. Psychometric test performance. *Archives of Neurology, 46,* 383-386.

Stuss, D.T., Ely, P., Hugenholtz, H., Richard, M.T., LaRochelle, S., Poirier, C.A., & Bell, I. (1985). Subtle neuropsychological deficits in patients with good recovery after closed head injury. *Neurosurgery, 17*, 41-47.

Sweet, J.J., & Moberg, P.J. (1990). A survey of practices and beliefs among ABPP and non-ABPP clinical neuropsychologists. *The Clinical Neuropsychologist, 4*, 101-120.

Sweet, J.J., Moberg, P.J., & Westergaard, C.K. (1996). Five-year follow-up survey of practices and beliefs of clinical neuropsychologists. *The Clinical Neuropsychologist, 10*, 202-221.

Sweet, J.J., & Westergaard, C. (1997). Psychopathology and neuropsychological assessment. In G. Goldstein & T.M. Incagnoli (Eds.), *Contemporary approaches to neuropsychological assessment* (pp. 325-358). New York: Plenum Press.

Swiercinsky, D.P. (1987). Neuropsychological assessment of sensory and physically impaired patients. In L.C. Hartlage, M.J. Asken & J.L. Hornsby (Eds.), *Essentials of neuropsychological assessment* (pp. 196-212). New York, NY: Springer Publishing Company.

Tamkin, A.S., & Dolenz, J.J. (1990). Cognitive impairment in alcoholics. *Perceptual and Motor Skills, 70*, 816-818.

Taylor, E.M. (1959). *Psychological appraisal of children with cerebral deficits.* Cambridge, MA: Harvard University Press.

Thompson, L.L., Heaton, R.K., Matthews, C.G., & Grant, I. (1987). Comparison of preferred and nonpreferred hand performance on four neuropsychological motor tasks. *The Clinical Neuropsychologist, 1*, 324-334.

Trahan, D.E., & Larrabee, G.J. (1988). *Continuous Visual Memory Test.* Odessa, FL: Psychological Assessment Resources.

Troester, A.I., Fields, J.A., Paolo, A.M., & Koller, W.C. (1996). Performance of individuals with Parkinson's disease on the vocabulary and information subtests of the WAIS-R as a neuropsychological instrument. *Journal of Clinical Geropsychology, 2*, 215-223.

Uchiyama, C.L., D'Elia, L.F., Dellinger, A.M., Becker, J.T., Selnes, O.A., Wesch, J.E., Chen, B.B., Satz, P., van Gorp, W., & Miller, E.N. (1995). Alternate forms of the Auditory-Verbal Learning Test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology, 10*, 133-145.

Van Lancker, D., & Kempler, D. (1987). Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain and Language, 32,* 265-277.

Wang, P.L. (1977). Visual organization ability in brain-damaged adults. *Perceptual and Motor Skills, 45,* 723-728.

Warren, D.H. (1992). Cognitive development in visually impaired children. In S.J. Segalowitz & I. Rapin (Eds.), *Handbook of neuropsychology, Vol. 7: Child neuropsychology* (pp. 15-31). Amsterdam: Elsevier.

Warrington, E.K. (1984). *Recognition Memory Test.* Windsor, United Kingdom: NFER-Nelson.

Warrington, E.K., James, M., & Maciejewski, C. (1986). The WAIS as a lateralizing and localizing diagnostic instrument: A study of 656 patients with unilateral cerebral lesions. *Neuropsychologia, 24*, 223-239.

Webber, A.M. (1988). A new clinical measure of attention: The Attentional Capacity Test. *Neuropsychology, 2,* 59-71.

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

Wechsler, D. (1981). *WAIS-R manual.* New York, NY: The Psychological Corporation.

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

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997a). *WAIS-III, WMS-III technical manual.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *WAIS-III: Administration and scoring manual.* San Antonio, TX: The Psychological Corporation.

Wedding, D., & Faust, D. (1989). Clinical judgement and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4,* 233-265.

Wetzel, L., & Boll, T.J. (1987). *Short Category Test, booklet format.* Los Angeles: Western Psychological Services.

Wetzel, L., & Murphy, S.G. (1991). Validity of the use of a discontinue rule and evaluation of discriminability of the Hooper Visual Organization Test. *Neuropsychology, 5,* 119-122.

Wilde, M.C., Boake, C., & Sherer, M. (1995). Do recognition–free recall discrepancies detect retrieval deficits in closed-head injury? An exploratory analysis with the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 17,* 849-855.

Wilkinson, G.S. (1993). *WRAT3 Administration Manual.* Delaware: Wide Range.

Wilmot, C.B., Cope, D.N., Hall, K.M., & Acker, M. (1985). Occult head injury: Its incidence in spinal cord injury. *Archives of Physical Medicine and Rehabilitation, 66,* 227-231

Wilson, B., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome.* Bury St. Edmunds, England: Thames Valley Test Company.

Wilson, B., Cockburn, J., & Baddeley, A. (1985). *The Rivermead Behavioural Memory Test.* Bury St. Edmunds, England: Thames Valley Test Co.

Wilson, B.C. (1987). Neuropsychological assessment. In V.B. Van Hasselt & M. Hersen (Eds.), *Psychological evaluation of the developmentally and physically disabled* (pp. 69-89). New York: Plenum Press.

Wilson, R.S., & Kaszniak, A.W. (1986). Longitudinal changes: Progressive idiopathic dementia. In L.W. Poon (Ed.), *Handbook for clinical memory assessment of older adults* (pp. 285-293). Washington, DC: American Psychological Association.

Wood, J.M., Tataryn, D.J., & Gorsuch, R.L. (1999). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*, 354-365.

Woodward, A.C. (1982). The Hooper Visual Organization Test: A case against its use in neuropsychological assessment. *Journal of Consulting and Clinical Psychology, 50*, 286-288.

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

Youngjohn, J.R., Larrabee, G.J., & Crook, T.H., III. (1992). Test-retest reliability of computerized, everyday memory measures and traditional memory tests. *The Clinical Neuropsychologist, 6*, 276-286.

APPENDICES

APPENDIX A. PUBLICATIONS ARISING FROM THE RESEARCH

Neuropsychological Assessment Services for People with Physical and Sensory Disabilities

Duncan R. Babbage and Janet M. Leathem
School of Psychology,
Massey University

Neuropsychological assessment is the examination of the cognitive, affective and psychosocial functioning of an individual. After a traumatic brain injury, neuropsychological assessment is useful both for tracking recovery in the acute stage and, if required, for planning long-term rehabilitation and future directions in later stages. Neuropsychological assessment measures typically require either verbal or fine motor responses. In individuals with physical and sensory disabilities performance on these measures can therefore be impaired, regardless of their level of cognitive functioning. This paper presents the current state of the field in assessing these individuals. At this time, while strategies exist for such assessments, much further work is required. The current research programme that aims to develop more robust assessment techniques is outlined.

The aims of this paper are fourfold. Firstly, this paper will provide an overview of just what neuropsychological assessment is. Secondly, it will examine some of the things that can make conducting a neuropsychological assessment difficult. The main focus of this presentation will be on one of these things–the difficulties in assessing people with physical and sensory disabilities. Finally, our current research programme on this topic will be outlined.

Some readers may be well aware of what neuropsychology is about, having worked in teams with neuropsychologists, or having had some experience of neuropsychological assessment. However, for those who are not familiar with neuropsychology, a brief overview will be provided. Three questions will be answered about neuropsychology: Who does it? What is it? and Why do it?

Neuropsychological assessment is conducted by Registered Psychologists with specialised training in neuropsychology. Assessments may be conducted in hospitals, clinics, such as the Psychology Clinic at Massey University, or by practitioners in private practice. With assessments following head injury in New Zealand, assessments are usually funded by ACC.

One textbook defines clinical neuropsychology as "an applied science concerned with the behavioural expression of brain dysfunction" (Lezak, 1995, p. 7). Neuropsychologists are interested in examining a person's ability to display various aspects of "intelligent behaviour", and investigate a person's areas of cognitive strength and weakness. They examine functioning using tests which require the person being assessed to answer questions, do puzzles, and so on.

An important aspect of these tests is that they have normative information, which means that a person's performance can be compared to other healthy people similar to them, often matched to their age, ethnicity, gender, and level of education. In this way the areas in which a person is having particular difficulty can be identified. The tests used have been extensively researched, and have been shown to really measure what they are designed measure, and to do so reliably.

Some of the areas that are measured in neuropsychological assessment include attention & concentration, motor functioning, perception, construction, language functioning, speed of information processing, memory and learning, and executive functions (the ability to do things like plan and monitor our own behaviour). Emotional and interpersonal functioning are also assessed. For individuals with head injuries those areas most commonly affected tend to be memory and learning, executive functions, information processing speed, and emotional and interpersonal functioning.

In the past neuropsychological assessment has been used to diagnose the presence of brain pathology (e.g., brain tumours, stoke, head injuries). Improved neuroimaging, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) has changed this role. Where neuroimaging reveals damage to the brain, neuropsychological assessment can determine the functional effects of that–how the damage affects a person's ability to do things. Additionally, individuals may have a clear CT scan and still be experiencing significant difficulty. Therefore, neuropsychological assessment may often be useful even in cases without specific neuroradiological findings.

When planning rehabilitation the more information that can be put on the table for consideration the better. As well as knowing that a person has a memory difficulty it is important to establish the actual nature of the deficits, so any rehabilitation planned can play to their areas of strength. Neuropsychological assessment can establish the nature of such difficulties. Planning rehabilitation from a position of knowledge and awareness will ensure the best outcome for the client.

However, there are some factors that complicate the assessment process. For example, a client may be so depressed following their injury that they are unable to complete the tasks required for an assessment. An individual may have a condition such as schizophrenia in addition to a brain injury, which will likewise interfere with formal neuropsychological assessment. Alternatively, some individuals may be unmotivated if they feel the assessment is unnecessary, and thus not give their best performance. Contextual factors can play a part: a quiet office is preferable and if a noisy ward is the only available site for assessment then results may be compromised. Medical factors, such as the medications a person is on, can alter levels of attention and concentration abilities, and thus affect testing. This applies likewise to alcohol and non-prescription drugs used.

Finally, disability factors can impede assessment. Not all disabilities affect neuropsychological testing. Since the tests involve things like answering questions, drawing things, and writing, a person paralysed from the waist down could complete all the tests without difficulty. In some cases, the disability may also mean some assessment is not required. For example, if a person is completely blind, there is no need to be able to assess their ability to perceive visual information, since this is an ability they can no longer use in any functional sense.

However, there are some disabilities which do directly impede the neuropsychological assessment process. Test materials usually are presented visually, verbally, or both, and responses are usually made by drawing, writing, manipulating objects, or speaking. Where test materials cannot be presented in the standard way, or where a person cannot physically make the required response, standard assessment may be blocked.

Where standard neuropsychological assessment is disrupted, there are some alternatives, including using alternative assessment measures, which use different modes of presentation or response. Additional use of behavioural data, such as observing a person at home, work or at school is also useful. Greater use of collateral information, from relatives, teachers, co-workers, and friends may also be used.

However, it is our view that further work is needed to improve the neuropsychological assessment services that can be offered to individuals with physical and sensory disabilities. Therefore we are conducting a research programme in this area. The aim of this research is to improve standard neuropsychological assessment procedures so they can be given to clients with a wide range of disabilities.

The first step in this programme was a review of 287 people referred to the Psychology Clinic, Massey University, for neuropsychological assessment following brain injury between 1988 and 1997 (Babbage & Leathem, 1997). We found that 8% of this sample had physical and/or sensory disabilities that interfered with the standard assessment process. This sample may not be representative of the general population of people with head injuries. These tended to be individuals who had experienced severe traumatic brain injuries, who had often had their accident some years before, and were still experiencing difficulties. It may also be that these "hard-to-assess" clients, with physical and sensory disabilities, may be referred more frequently to our clinic as it is an area of specialisation. However, this is nonetheless a significant group of people, and it is important that we are able to provide the most appropriate services that we can.

Following from this review, a survey was conducted of a group of neuropsychological experts, both in New Zealand and overseas (Babbage & Leathem, 1998). In this survey we examined the current assessment practices of neuropsychologists when working with clients with physical and sensory disabilities. We found essentially that while many of these individuals had strategies for working with such people, there were no guidelines available to direct the assessment of these "hard-to-assess" clients.

As the next step in this research programme, we are developing a protocol to assess the range of cognitive functioning using communication methods available to individuals with severe physical and sensory disabilities. In order to do this we are finding some existing neuropsychological measures that can be completed without having to be able to speak, write or draw. Also, we are adapting existing measures to be suitable for these individuals.

These measures will be administered to a group of individuals who have experienced head injury, and who also have significant physical and sensory disabilities. Ultimately we want to be able to assess a person with this level of disability and know whether they are experiencing any cognitive difficulty. Therefore, we need to know how people who have not had a head injury would do on our tests. Therefore, we will also give these tests to a group of individuals with similar disabilities, who have not experienced a head injury.

In addition to these two groups, we will also give these tests to a group of healthy individuals who do not have any disabilities, and who have not had a head injury. This will give us information about the effects of head injury, and of physical and sensory disabilities, on ability to complete measures, as we will be able to compare the performance of these healthy individuals to those in the other two groups. Healthy individuals will be required to complete the tasks using response mechanisms that would be available if they had the same level of disability as the other groups–that is, they may be required to use an alphabet board, or to communicate using eye movements. We will also give them standard versions of the tests, so we can see how our adaptations compare with these existing measures.

With these three groups we can examine the extent to which these measures are affected by physical and sensory disabilities, and the extent to which they are able to discriminate between people who have experienced head injuries, and those who have not. This is important because the aim of developing these measures is to have tests which are sensitive to brain damage, but are not sensitive to the effects of physical and sensory disabilities.

In the next stage of our research we will be specifically focussing on people with one particular set of disabilities. They will be people who have had a traumatic brain injury, and in addition have a motor disability which affects their fine motor control—their ability to write and draw—and have an expressive speech disability—that is, are unable to communicate effectively with speech. We will be working with people with whom communication has been reliably established in some way—this may be though the use of an alphabet board, or some other method.

In this phase of the research we are clearly focussing on a very specific group. This group was selected as it was the largest of the groups of people with disabilities that were identified in the prevalence study. However, the knowledge produced by this research will be applicable to other groups also.

The status of this research is that the adapted protocol is currently being developed, and will soon be being tested with the first clients at the Psychology Clinic at Massey University. While we do not have the money available to fly to Invercargil, in the future it is anticipated that some people who are in our target group may also be seen in other major metropolitan areas, particularly in the North Island.

This should not be seen as a back-door way of getting a neuropsychological assessment. As we are currently developing the protocol and have not yet established its reliability and validity, we cannot make reports on specific cases. Nonetheless, if you work in a service which might be a suitable place for us to recruit individuals to participate in this research, or you know of someone who might be suitable, we would be pleased to send you out some further information about the research. The current timeframe is that this information is likely to be sent out in November or December of 1998, and assessments will be being conducted in late 1998 and into 1999. If you would be interested in this, please make yourself known to us.

We can be contacted at the Psychology Clinic, Massey University, Private Bag 11-222, Palmerston North. Our phone number is 06- 350 5196, or you can email us at <D.R.Babbage@massey.ac.nz>.

References

Babbage, D.R., & Leathem, J.L. (1997, September). *Neuropsychological assessment: Difficulties associated with "hard-to-assess" clients.* Paper presented at the New Zealand Psychological Society Conference, Palmerston North, New Zealand.

Babbage, D.R., & Leathem, J.L. (1998, August). *Neuropsychological assessment of "hard-to-assess" clients: Practitioner survey.* Paper presented at the New Zealand Psychological Society / New Zealand College of Clinical Psychologists Conference, Wellington, New Zealand.

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

Assessment practices of neuropsychologists for individuals with physical and sensory disabilities

Duncan R. Babbage & Janet M. Leathem
School of Psychology
Massey University
Palmerston North, New Zealand

Background

Neuropsychological assessment with individuals with physical disabilities presents a significant challenge. A review of the literature reveals few useful guidelines for working with these very hard to assess individuals, especially where an individual is impaired in more than one modality.

Aims

To examine the assessment practices of a group of neuropsychologists when working with individuals who have physical and sensory disabilities. This exploratory study aimed to highlight issues to inform both clinical practice and further research.

Method

As this research was exploratory a random sample was not attempted. Questionnaires were initially distributed at the INS conference, Honolulu, in January 1998. A further 23 were directly mailed to neuropsychologists. In all 34 questionnaires were returned (52%).

Questionnaire

Examined the approach respondents would take to assessment, particularly that of an individual described in a presented case vignette. The vignette described the fictitious case of an individual who sustained a closed head injury in a motor vehicle accident.

Case Vignette

"L.M. is a 31 year old female who suffered a closed head injury in a motor vehicle accident where the car she was driving collided with a lamp post. L.M. was admitted to hospital unconscious, and remained so for 16 hours. Post-traumatic amnesia was estimated at 36 hours, although the effects of medications she was receiving made it difficult to establish this precisely. Since her injury L.M. has been unable to speak, although she appears to understand spoken English. Spinal injuries left her paralyzed from the waist down, with only gross motor control currently regained in her arms. Reliable communication has been established with pointing and the use of an alphabet board with assistance. You have been referred this patient to assess the degree of cognitive impairment, in order that decisions can be made regarding rehabilitation and L.M.'s future directions."

Participants

Participants were neuropsychologists from the United States (n=21), New Zealand (n=9), Canada (n=2) and Australia (n=2). 59% were in full time employment in neuropsychology, 32% part time, the remainder completing training.

Poster presented at the 3rd World Congress on Brain Injury, Quebec City, Quebec, Canada, June 15th 1999.

Results – *Referral*

Figure 1. Proportion of respondents who would accept referral of L.M.

Difficulties

Issues identified by respondents:
Assessment was not possible solely through their standard techniques and procedures.
Motor control and language key blocks to standard techniques.
Motor slowing would mean direct comparison to norms inappropriate.

First Steps

Assessment of communication abilities, including examination of response reliability.
Examine methods of response other than alphabet board, particularly simple Yes/No.
Effects of medications, attention, fatigue, and frustration may be exacerbated.

Strategies

Adapted assessment measures, particularly to Yes/No format.
Increase use of collateral information, from family, caregivers, and other professionals.
Use of behavior rating scales and observation of patient in their environment.

Measures

The following measures were endorsed for use with L.M.:

Peabody Picture Vocabulary Test
Vineland Adaptive Behavior Scales
Sensory-Perceptual Examination
Hooper Visual Organization Test
Raven's Progressive Matrices
Beck Depression Inventory
Seashore Rhythm Test
Grip Strength
Facial Recognition Test
Benton Line Orientation

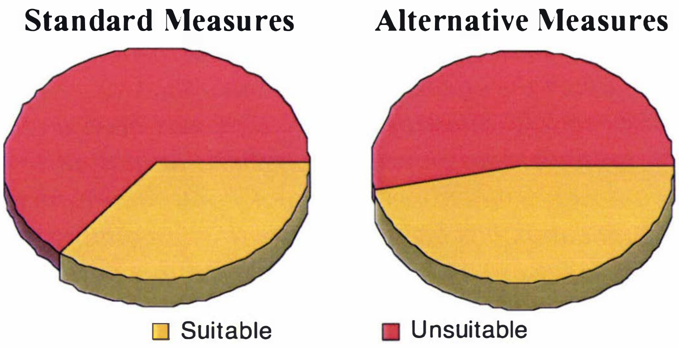


Figure 2. Proportion of participants' standard and alternative measures which were seen as suitable for use with L.M.

Discussion

All participants indicated changes to standard practice would be required. From the measures identified, memory & learning, concept formation & reasoning, and executive functions stand out as areas requiring further assessment.

Assessment of cognitive functioning in individuals with expressive disabilities in addition to traumatic brain injuries.

Duncan R. Babbage & Janet M. Leathem
Massey University
Palmerston North, New Zealand

Background to the Research Project

Clinical psychologists conduct neuropsychological assessments in order to determine the strengths and weaknesses of an individual after a brain injury, and to aid in rehabilitation planning. Where individuals present with physical or sensory disabilities, neuropsychological assessment is difficult, even for the most well equipped settings, as few clinically useful guidelines exist. Clinicians may often make ad-hoc adaptations of measures, and interpret the data qualitatively. Systematic assessment strategies for individuals with these kinds of disabilities would enable more comprehensive assessment, and thus better targeted rehabilitation efforts.

Objective of this Paper

This paper presents part of a project which aims to provide information on standardising cognitive assessment of individuals with severe expressive disabilities — who cannot speak, write, draw, or manipulate test materials — but with whom reliable communication has been established[1]. Such individuals may typically communicate using an alphabet board or computer device. An overview will be presented of the assessment battery developed in this research. Process issues involved in cognitive assessment of individuals with this level of disability will be discussed, and the paper will present initial qualitative information about the usefulness of the measures in the research battery. Finally, strategies for the interpretation of non-standard neuropsychological assessment will be discussed.

Research Measures

A battery of tests that assesses a broad range of cognitive functions in individuals with expressive disabilities was assembled. Measures in the research battery were selected and/or adapted to meet the following criteria:

- All responses made with a simple pointing response.
- No verbal communication is required.
- No manipulation of test materials is required.
- No component of scoring is based on time.

Many of these tests were adapted from measures which clinicians will already have access to. The measures used in this research are presented in Table 1. This battery is currently being administered to individuals with brain injuries who have severe expressive disabilities, and to comparison groups of healthy individuals and individuals who have only had severe traumatic brain injuries.

Table 1. Measures adapted for severely disabled individuals.

Domain Assessed	Measure
Orientation	Good Samaritan Hospital Orientation Test
Sustained attention	Graded Attentional Test
Short-term memory span	Digit Span Forwards
Working memory	Digit Span Backwards
Verbal memory	Verbal Paired Associates (WMS-III)
Verbal memory	Logical Memory (WMS-III)
Nonverbal memory	Memory for Faces (WMS-III)
Nonverbal memory	Family Pictures (WMS-III)
Language – sequential vs. holistic	Familiar and Novel Language Comprehension Test
Language – confrontational naming	Boston Naming Test
Language – comprehension	Auditory Reception (ITPA)
Nonverbal reasoning	Matrix Reasoning (WAIS-III)
Visuoperception	Hooper Visual Organization Test
Executive functions	Match and Shift Categories Test

---

[1] This research is supported by a Health Research Council of New Zealand Clinical Research Limited Budget Grant.
Paper presented at the Australasian Winter Conference on Brain Research, August 1999, Queenstown, New Zealand.

Adaptations Made to Research Measures

Table 2 presents the adaptations that were made to the measures selected. Two measures were used in their standard administration. For the majority of the remaining measures, administration was changed only in that simple responses were selected from cards (e.g., Yes/No, or pointing to family members' names), and other responses were spelt out using an alphabet board (or their communication device).

However, two measures were more significantly adapted. For Logical Memory participants were immediately administered the recognition questions using a Yes/No card, without a free recall trial. Additionally, Story A recognition was administered immediately only, and Story B recognition was administered only after a delay. Secondly, the Verbal Paired Associates task was administered by selecting responses from a multiple-choice card, which provided participants with a 1 in 8 selection as opposed to the standard free recall.

Table 2. Adaptations Made to Research Measures

Standard Administration
Matrix Reasoning
Familiar and Novel Language Comprehension Test
Yes/No Response Card
Good Samaritan Hospital Orientation Test
Auditory Reception
Memory for Faces
Logical Memory
Multiple-Choice Response Card
Family Pictures
Verbal Paired Associates
Number Board Response Card
Digit Span Forwards
Digit Span Backwards
Alphabet Board Response Card
Boston Naming Test
Hooper Visual Organization Test
Created for this Research Project
Graded Attentional Test
Match and Shift Categories test

Two measures were specifically developed for this project, although both were based partly on previous measures. The Graded Attentional Test (GAT) was designed to examine sustained attention and was based on the broad principles of the Attentional Capacity Test (ACT). The new measure contains fewer trials, fewer stimuli, and fewer levels of difficulty than the ACT, making it more suitable for the requirements of this research. Like the ACT, the GAT requires only a single digit response for each trial, which for the GAT is selected on a number board.

The Match and Shift Categories test (MASC) is broadly similar to a number of category selection tasks used to examine the ability to perceive the rules underlying a situation, to maintain an established perceptual set, and measuring perseveration in the face of changing task demands. The MASC requires participants only to indicate a selection between four choices by pointing.

For full details of all adaptations of test materials and administration instructions, interested parties should contact the authors.

Process Issues in Assessment

For many clinicians, the following section will be familiar territory. Nevertheless, these points are presented as a 'checklist' of factors to consider, recognising that in highly complex cases it can at times be difficult to see the wood for the trees:

Prepare Well

While in such cases not all eventualities can be predicted, having a selection of assessment options available before beginning assessment will be useful. Many adaptations, such as those used in this project, only require simple additional materials. However, these are best thought out in advance, rather than produced ad hoc during an assessment.

Establish Reliable Communication

Before any form of neuropsychological assessment can be undertaken, either standardised or ad hoc, a reliable communication technique must be identified. In some cases, this will be well developed, such as the regular use of a Lightwriter or alphabet board; in other cases while functional communication may be often employed, a standard and demonstrably reliable method may be difficult. Initial tasks should clearly demonstrate the reliability of the response mechanism that is to be used in the assessment.

Plan for Short Sessions

Traumatic brain injury is, of course, associated with increased levels of fatigue. Severe physical disabilities add an additional level of both fatigue and frustration to the assessment process. However, individuals with this level of expressive disability are often highly motivated to participate in the assessment process, partly as the assessor demonstrates such a clear interest in communicating with them, an understanding what they are trying to say. Of necessity, assessments will have to be short, although with sufficient sessions considerable information can be gleaned from the process.

Involve Caregivers and Significant Others

Very few individuals who are “non-verbal” use no vocal communication at all. Often caregivers and particularly family members will be able to interpret both verbal and gestural communication, allowing rapport to be built more quickly, and reducing the frustration experienced by both clinician and client. While any assistance in interpreting responses must clearly be carefully handled in a formal testing situation, some literature indicates that family members may be able to reliably interpret communications which professionals view as unintelligible (Davies & Mehan, 1988).

Be Flexible

While standardised assessment is the ideal goal, as always assessors need to be ready to change the task requirements to be suitable for the individual being assessed. While this raises difficulties with interpreting performance relative to normative groups, there are some ways of interpreting such ad hoc data systematically. Fortunately, these methods are particularly well suited to the kinds of adaptations typically made for people with expressive disabilities. [See “Interpretation of Non-Standard Assessment” below].

Qualitative Evaluation of the Measures

While early results suggest that many of the adapted measures described in this paper may be useful instruments for the assessment of cognitive functioning in individuals with expressive disabilities, formal psychometric evaluation is still currently underway. However, a number of qualitative observations can be made, based on the assessments conducted so far, which may aid clinicians in conducting assessments of this sort.

You Cannot Assess All of the Domains, All of the Time

In selecting the measures for this project, some domains of cognitive functioning normally assessed stood out as extremely difficult to assess in such individuals, and ultimately no measure was included of these domains in the protocol. This included visuoconstruction, information-processing speed, motor functioning, and verbal memory tasks involving free-recall. Some of these domains may not need to be assessed in a cognitive assessment of a client with severe expressive disabilities. Motor difficulties may be self-evident at a gross level, and best assessed by a Physiotherapist or Occupational Therapist at a specific level. Motor tasks such as the Finger Tapping Test are typically conducted in a neuropsychological assessment to aid in the interpretation of other tests that have an incidental motor, or timed, component. In the current assessments, neither of these considerations applies.

Due to the emphasis on speed in most information processing tasks, no suitable measure of this domain could be identified in the literature. A few possible measures were conceived by the authors, which involved processing information at speed but did not require an immediate, or speeded, response. However, these were felt to be too cumbersome and open to confound to be suitable for the current project. Nonetheless, in some high functioning individuals with severe expressive disabilities, it is conceivable that this might be an area of investigation. (For example, an assessor might be asked whether a particular individual take in the information in lectures, which are typically presented at high speed, or whether the lectures should be taped, and presented for the client at a reduced rate?). Similarly, while construction may not be possible for individuals with severe physical disabilities, visuospatial abilities typically tapped by such measures may be of considerable importance in decisions regarding suitable wheelchair control mechanisms. In cases such as these described, however, contextualised functional analysis of the specific abilities and needs of the individual are likely to prove of more benefit in answering these questions.

Less Is More

Standard neuropsychological assessment measures typically require participants to produce as much as they can by themselves initially, and after this may then provide a cued or recognition trial. Discrepancy between spontaneously produced and cued responses can indicate the specific nature of any deficit, and are employed in the assessment of diverse areas such as language and memory. However, when assessing individuals with expressive disabilities, initial results of this project suggest that, in at least some cases, less may be more. While freely produced information (e.g., spelling out sentences on an alphabet board) most closely resembles standard task requirements, there may be too many factors confounded with poor performance.

One example is motivation to communicate all available information. For example, on the Family Pictures subtest of the Wechsler Memory Scale-III participants are score on the volume of the (correct) information recalled. When instructed not to speak their answers, and asked to spell out what family members were doing in the pictures, participants in both the healthy and traumatically brain injured comparison groups have been seen to spell a relatively brief answer — e.g., "Sitting" — and follow it verbally with "at the wooden picnic table, shielding themselves from the dog who is splashing water on them; and there was three bottles on the table, mustard, tomato, and something else". When asked again to spell out their answers, some participants in further questions provided only short answers. Clearly, even in normative groups, participants are not motivated to provide all the information they may have.

For individuals with expressive disabilities, there clearly is no easy way to gauge if they have this additional information. Additionally, if such an individual has been in an environment where caregivers are too busy for protracted communication, individuals may have been reinforced for communicating in brief, clipped responses. They thus may be even less likely to give full answers, even where they have this level of information available.

Tests could be structured to provide prompting ("What else can you remember?" or "Where was he sitting?"), partially compensating for such issues. It is possible, however, that a more robust solution may be to develop more tests which focus on multi-choice responses. In such a context, a participant who gives any response at all has given the most full response possible. Motivational factors, short of malingering, would not play such a significant role in scores on these tests.

Interpretation of Non-Standard Assessment

While where possible measures will be administered under standardised conditions, in some cases non-standard assessment may be necessary. Interpretation of results obtained from such assessments is still possible.

Lowest Estimate of Actual Ability

In many cases, it may reasonably be argued that ad hoc adaptations of measures have made the tasks more demanding. For example, it is more difficult to hold the content of a story in memory while spelling it out on an alphabet board, than to repeat it verbally. In such a case, where a person's score, using the adapted administration, falls into a particular band in the normative group, it can be concluded that their true ability must be *at least* that high. For example, if on the Logical Memory test a person falls at the mean for the normative group, using the administration described above, one can reasonably conclude that their true level of ability is at least average in this area. However, in such a context no parallel interpretation can be made for low scores — there are too many uncontrolled factors and results must be considered at best highly equivocal.

Statistical vs. Normative Analysis
Depending on the nature of the referral question, in some cases establishing any level of ability in an individual with severe expressive disabilities may be an important step. For instance, indication of a reliable yes/no response, and of comprehension of complex language may be two very important steps in determining competence in a legal context. In such cases while normative data should be referred to where possible, the issue is not so much a person's position in the general population, but rather their basic ability to understand, remember, process and communicate information.

In such cases, statistical procedures may be used to interpret performance of non-standard assessment. Such procedures, basically similar to those used in malingering tests, examine an individual's performance relative to chance. In any case where an individual is selecting between a limited range of alternatives it is a simple matter to determine the average performance of someone who is responding in a random fashion, and the number of correct responses required to be confident (at a certain level of statistical significance) that a person is responding above chance level. Again, while such performance may represent functioning at levels considerably below their premorbid functioning, it can provide quantification of areas of strength and weakness. As stated earlier, many of the measures suitable (with adaptations where required) for individuals with expressive disabilities meet the criteria above — they require an individual to respond by selecting between a fixed number of choices. Again, while some literature has been published on this type of interpretation (e.g., Sinnett & Holen, 1995), this is an area which might warrant further investigation.

Implications of the Findings
While individuals with expressive disabilities are challenging for neuropsychologists, a number of findings from this research may be of assistance in conducting an adequate assessment. First, while few clinicians see many clients with this level of disability, a number of simple adaptations to standard instruments on hand may be useful.

Secondly, identifying which tests will provide the least confounded, and therefore most useful, information is important. Some tests commonly used by neuropsychologists, such as the Matrix Reasoning subtest of the WAIS-III, can be used without any adaptation.

Finally, where standardised assessment is not possible, there are still objective ways of analysing the data obtained. Such analysis may not provide the same level of detail about functioning which neuropsychologists would normally expect, but still provide valuable, objective, data to guide clinical judgement in conducting an assessment.

References

Davies, P., & Mehan, H. (1988). Professional and family understanding of impaired communication. *British Journal of Disorders of Communication, 23,* 141-151.

Sinnett, E.R., & Holen, M.C. (1995). Chance performance in the absence of norms for neuropsychological tests: A clinical note. *Journal of Clinical Psychology, 51* (3), 400-402.

APPENDIX B. NEUROPSYCHOLOGICAL PRACTICE SURVEY

Neuropsychological Practice Survey

Duncan R. Babbage & Janet M. Leathem
Massey University, Palmerston North, New Zealand

Demographics

1. Gender: F M 2. Age:

3. Highest degree: Ph.D Psy.D. Ed.D Diplomate Other (Please indicate):

4. Country in which you are licensed to practice (if USA, State in which you are licensed):

Clinical Practice

5. Do you personally interview patients during evaluations? Yes No

6. Do you use a technician/psychometrician or other assistant to collect test data from your patients? Yes No

7. Do you personally observe patients during any or all of the administration of neuropsychological tests? Yes No

8. Indicate the *total* amount of time that you spend doing a single neuropsychological evaluation (include test administration *and* scoring *and* interpretation): _______ hours per evaluation.

9. Please indicate your *primary* philosophical approach toward test selection in neuropsychological assessment (even if you vary at times from this approach) (check one only):

____ *Flexible* (i.e., based upon the needs of an individual case; not uniform across patients).
____ *Flexible Battery* (i.e., variable but routine groupings of tests for different *types* of patients, such as head injury, alcoholism, elderly, etc.).
____ *Standardized Battery* (e.g., Halstead-Reitan, Luria-Nebraska, Benton, or other routine grouping of tests that is uniform across patients).

10. Which best describes your current work status in clinical neuropsychology? (check one only)

____ Full-time active employment
____ Training: fellow, intern, student, etc.
____ Other: please describe:
____ Part-time employment
____ Retired, but work occasionally

11. What percentage of your total clinical time is spent in the following?

____% Neuropsychological assessment
____% Neuropsychological treatment (direct contact and involvement with patients)
____% Neuropsychological consultation (organizations, attorneys, physicians, etc.)

12. Over all your employment settings, what percentage of your clinical time is spent working with patients who have suffered a traumatic brain injury? ____%

Core Assessment Measures

13. From the list of measures on the following page, please check in the first box next to each measure that you would PRIMARILY use for assessment of an adult patient with a CLOSED HEAD INJURY in the post acute stage. (If you use only part of a standardized battery, check those measures which you use). Please add any measures which are not listed, at the bottom of the page.

Secondary Measures

14. Please now put a check in the second box beside those measures that you would use as secondary measures if one of the core assessment measures proved to be unsuitable, or if you wished to further investigate a domain of functioning.

(If you would not use additional measures, put a check/tick in this box: ☐)

Halstead-Reitan Battery			
Aphasia Screening Test (Rev)			
Category Test			
Finger Tapping Test			
Fingertip Number-Writing Percep.			
Grip Strength/Hand Dynamometer			
Seashore Rhythm Test			
Sensory-Percpetual Examination			
Speech Sounds Perception Test			
Tactile Finger Recognition			
Tactual Performance Test (Sequin Formboard)			
Trail Making Test			
Luria-Nebraska Battery			
Luria-Christensen Battery			
Luria Approach			
Non-Standardized Battery			
WAIS			
WAIS-R			
WAIS-R-NI			
WAIS-III			
Arithmetic			
Block Design			
Comprehension			
Digit Span			
Digit Symbol Coding			
Digit Symbol Incidental Learning			
Information			
Letter Number Sequencing			
Matrix Reasoning			
Object Assembly			
Picture Arrangement			
Picture Completion			
Similarities			
Symbol Search			
Vocabulary			
WMS			
WMS-R			
WMS-III			
Family Pictures			
Letter Number Sequencing			
Logical Memory			
Memory For Faces			

Mental Control			
Spatial Span			
Verbal Paired Associates			
Visual Reproduction			
Word List			
MMPI			
MMPI-II			
Aphasia Screening Exam			
Babcock Levy			
Beck Depression Inventory			
Beery Visual-Motor Integration			
Bender-Gestalt			
Benton Line Orientation			
Benton Visual Retention Test			
Bicycle Drawing			
BNI Screen For Higher Cerebral Funct.			
Boston Aphasia Exam			
Boston Naming Test			
California Verbal Learning Test			
Cancellation Test			
Cognitive Estimate Questions			
Consonant Trigrams			
Controlled Oral Word Association Test			
Cookie Theft Picture			
Cowboy Story			
Dementia Rating Scale			
Denman Neuropsychology Memory Scale			
Design Fluency			
Detroit Test			
Draw a Clock-House-Tree-Person			
Everyday Memory Questionnaire			
Facial Recognition Test			
Galverston Orientation Amnesia Test			
Gray Oral Reading Test			
Grooved Pegboard			
Hand Preference Test			
Harris Tests of Lateral Dominance			
Hidden Figures Test			
Hooper Visual Organization Test			
Illinois Test Of Psycholinguistic Ability			
Katz Adjustment Scale			
Knox Cube Test			
Lateral Dominance Examination			

Mini Mental State			
Montreal Neurological Institute Battery			
Motor Impersistence Test			
Multilingual Aphasia Exam			
National Adult Reading Test (NART)			
Neurosensory Center Comprehensive Examination for Aphasia (NCCEA)			
Paced Auditory Serial Addition Test (PASAT)			
Peabody Picture Vocabulary Test			
Porteus Mazes			
Purdue Pegboard			
Quality Extinction Test			
Randt Memory Test			
Raven's Progressive Matrices			
Rey Auditory Verbal Learning Test			
Rey-Osterreith Complex Figure Test			
Rivermead Behavioural Memory Test			
Road-Map Test			
Rorschach			
Selective Reminding Scale (Bushke)			
Sentence Repetition Test			
Serial Digit Learning			
Shipley Institute Of Living Scale			
Stroop Color-Word Test			
Symbol Digit Modalities Test			
Symptom Check List-90 (SCL-90)			
Tactile Form Recognition			
Temporal Orientation Test			
Thematic Apperception Test (TAT)			
Three-Dimensional Praxis			
Tinkertoy Test			
Token Test			
Tower of Hanoi			
Twenty Questions Test			
Vigotsky Test			
Vineland Adaptive Behavior Scales			
Western Aphasia Battery			
Wide Range Achievement Test (WRAT)			
Wisconsin Card Sort			
Woodcock Johnson Psychoeducational Battery			
Woodcock Reading			

Other measures you would use:

Case Description

Please read the case description below. Imagine you had been referred this patient for assessment of cognitive functioning.

> L.M. is a 31 year old female who suffered a closed head injury in a motor vehicle accident where the car she was driving collided with a lamp post. L.M. was admitted to hospital unconscious, and remained so for 16 hours. Post-traumatic amnesia was estimated at 36 hours, although the effects of medications she was receiving made it difficult to establish this precisely. Since her injury L.M. has been unable to speak, although she appears to understand spoken English. Spinal injuries left her paralyzed from the waist down, with only gross motor control currently regained in her arms. Reliable communication has been established with pointing and the use of an alphabet board with assistance. You have been referred this patient to assess the degree of cognitive impairment, in order that decisions can be made regarding rehabilitation and L.M.'s future directions.

15. Please note some of the difficulties (if any) that might arise assessing this patient with standard techniques:

16. Would you accept this referral for assessment? Yes No

17. On the list of measures (opposite) please CHECK in the third box any measures that you would use with this patient. At the same time, please put a CROSS in the third box next to measures which you would normally use, but which would be unsuitable with this patient due to her physical and sensory disabilities.

19. Please describe any particular assessment strategies you would use with this patient, and add any additional comments about the way you would handle this case:

20. If the assessment measures you wished to use were unsuitable due to the disabilities of a particular patient, what course of action would you take? (check / tick **one of the 6 options**):

Use standard measures as much as possible in the affected domains, and...
- ____ interpret the results *relative to* the original normative information.
- ____ interpret the results *clinically, without reference to* normative information.

Adapt existing measures, and...
- ____ interpret the results *relative to* the original normative information.
- ____ interpret the results *clinically, without reference to* normative information.

____ Exclude assessment of that cognitive domain.

____ Other (please state):

(contd. over)

Comprehensive Assessment

20. Please list the domains of cognitive functioning that, in your opinion, are important to assess in a comprehensive neuropsychological assessment:

21. In your view, is it important to always examine all of these areas in a neuropsychological assessment? (check/tick **one or more**):

____ Yes, regardless of context, patient, or referral questions
____ Depends on nature of referral questions
____ Depends on nature of patient presentation/difficulties
____ Depends on available resources in assessment context
____ Assessment focus can be directed solely by clinician
____ Other (please specify):

— Thank you for your time —

APPENDIX C. ADMINISTRATION INSTRUCTIONS

NEUROPSYCHOLOGICAL ASSESSMENT OF INDIVIDUALS WITH EXPRESSIVE DISABILITIES IN ADDITION TO TRAUMATIC BRAIN INJURIES

Administration Instructions

[Note: Instructions are in standard type, while text to be read out to participants is in *italics.* Underlining indicates where the instructions for this project deviate from the standard administration of a measure, and **do not** indicate emphasis to be placed on those words].

DEMOGRAPHICS A

This section should be completed before the assessment is begun.

INFORMATION SHEET AND CONSENT FORMS

Start stopwatch and record time on sheet. The instructions for this section are guidelines only; ensure participant understands all aspects of the information sheet, and that you have answered any questions that they have.

> *This information sheet tells you about this research. Please read it through and make sure you are happy with all the parts of the research. You can keep this sheet.*

Allow time to read sheet.

> *As it says in the information sheet, this research is about assessing people who have severe disabilities that mean they can not speak, write or draw. We are working to develop ways to assess memory, language and so on, that only require simple responses.*

If participant is in the TBI or Norm groups (i.e., not physically disabled), say:

> *For some of the things we'll do today I'll ask you to only use the kinds of responses that the participants with severe disabilities can use, such as selecting from some choices on a card like this* (show YES/NO card). *For the first part of the today, all the activities will be like this. For some of the later things we do, you will be able to answer in the way you normally would, such as by talking. I'll tell you before each part the way you can answer for that part. Do you have any questions about that?*

Answer any questions.

For all participants:

> *As we're doing the activities, I'll be recording the responses that you choose, mostly by circling them on this datasheet. You'll also see me writing down the time at various points... you don't have to worry, it's not a race. I just write down the time so that we can see whether people in the different groups take about the same amount of time to do all the activities.*

[TBI/Norm groups only: *In most of the activities...*] *There is no need to rush, you can go at whatever speed you like.* [TBI/Norm group only: *A couple of the activities we'll do near the end, I will ask you to go as fast as you can. I'll tell you when we get to these activities.*] *OK?*

Some of the activities we will do today you may find very easy. Some of them you may find really difficult. Don't be surprised by that. Some of the activities are deliberately designed to make you work hard, and anyone would have difficulty with them.

Are there any questions you want to ask me about the research?

Answer questions.

This consent form says that you understand what we'll be doing today, and you're willing to participate in this first assessment. Are you happy to participate in the study?

Ensure participant is familiar with the statements on the consent form, and that form is completed, before beginning.

RESPONSE RELIABILITY

This first activity shows you some of the cards you'll be choosing your answers from in some of the other activities. For each card, I'll say one of the things on the card, and your job is to point to that item. We do this with everyone to make sure that these cards are a good way to communicate. Are you ready?

Just before Coloured Shapes card, ask *"Are you colour-blind?"* Record in Demographics B.

DEMOGRAPHICS B

Which of these best describes your ethnic background. You can select more than one if you wish. [Show ethnicity card]

What is your age? [Number response card for disabled groups]

ED/TBI groups: *What job were you working in immediately before your accident?*

Norm group: *Are you currently working?* If yes, *What is your current job?*

If unemployed, record length of unemployment, and previous job held if any.

What is the highest level of formal education you completed?

If finished half way through Form 5, record Form 4, as this is the highest year completed. Also include polytech study, trade certificates etc. If any post-secondary education uncompleted record both completed secondary education and also how much of the course completed.

ED group: *Before your accident, were you right or left handed?*

TBI/Norm groups: *Are you right or left handed?*

If ambidextrous, *Did you write more often with your right or left hand?*

Record ID number, sex, group, date and location on p.1 at this point if not already done. Transfer age to GSHOT (p. 3) q3 and age+5 years for q4.

MEDICAL SCREEN

ED group:

> *I need to know a little bit about your accident, any other head injuries, and about any medications you take at the moment so we can understand the results of the research. Are you happy for* [name of caregiver] *to complete these questions for us later on?*

If response is Yes, ask questions to caregiver at the end of the assessment.

TBI/Norm groups:

> *I also need to find out a bit about your medical background. We are asking these questions because these things can change your performance on some of the tasks we will be doing today.*

Ask medical screen questions.

If any exclusionary criteria met, at end say:

> *One of the things that can affect how people do on these tasks is if they've had* (insert condition here) *in the past. For that reason we are excluding people who have had this. Therefore, we don't need to do any more. Thank you for your time today, and being willing to participate.*

Otherwise continue.

GOOD SAMARITAN HOSPITAL ORIENTATION TEST

Fill in blanks for questions to be asked before starting introduction.

> *In this activity, I'm going to as you a number of yes/no questions. I'd like you to select your response by pointing to one of the choices on this card* [indicate Yes/No card], *either Yes or No. If you make a mistake, point to "I made a mistake", and then make your choice again. Once you've made a choice, I'll ask the next question.*

BLOCK A

LOGICAL MEMORY A

I am going to read a short story to you. Listen carefully and try to remember it just the way that I say it, as close to the same words as you can remember. When I am through, I will ask you a series of questions about the story, which you can answer yes or no by selecting them from this card. You should answer as best you can remember even if you are not sure. Are you ready?

Anna Thompson of South Boston, employed as a cook in a school cafeteria, reported at the police station that she had been held up on State Street the night before and robbed of fifty-six dollars. She had four small children, the rent was due, and they had not eaten for two days. The police, touched by the woman's story, took up a collection for her.

Read questions.

At the end of the questions:

Now I am going to read another short story to you. As with the first story, try to remember it just the way I read it. Ready?

At 6:00 on Monday evening, Joe Garcia of San Francisco was watching television as he dressed to go out. A weather bulletin interrupted the program to warn that thunderstorms would move into the area within the next two to three hours and remain until morning. The announcer said the storm could bring hail and up to four inches of rain and cause the temperature to drop by fifteen degrees. Joe decided to stay home. He took off his coat and sat down to watch old movies.

I want you to remember as much of this story as you can because I will ask you to answer some questions about this story later.

FAMILIAR AND NOVEL LANGUAGE COMPREHENSION TEST

I'm going to say a sentence, and I'd like you to match one of these pictures to the sentence that I say. Some of the sentences are kind of funny. They're idioms. Do you know what an idiom is? Like "He has his head in the clouds".

Show Example Picture A

Which picture is best for "He has his head in the clouds?"

In example sections, give correct answer if participant selects wrong answer, and explain why answer is correct. Ensure the participant understands the concept of "literal" (i.e., novel sentences) and "non-literal" (i.e., familiar phases).

If choice is correct,

Yes, the boy is daydreaming, and it is true to say "He has his head in the clouds.

If choice is incorrect,

No, actually it's this one up here; the boy is daydreaming and so we would say that 'he has his head up in the clouds'.

How about these pictures. Which one is best for "The coast is clear"?

If correct, explain

Yes, the robber sees there's no one around, and so 'the coast is clear'. Is that what you thought too?

If incorrect, explain

No, it's this one; the robber sees there's no one around and so 'the coast is clear'.

If participant has missed both items, repeat the first two practice items, until it seems that the patient comprehends the idea of selecting pictures for these kinds of items.

Okay, now I'd like you to try to match the pictures for some sentences. I'll say each sentence once, so try to listen. I won't tell you the answers for these right way. Try to guess at an answer even if you're not sure. Are you ready?

Say each sentence clearly and naturally. Continue with Familiar Items. Do not provide feedback on response accuracy for test items.

At the end of the Familiar Items, say,

Now we will try regular sentences, not funny ones like before. Like this one, 'He takes his pets in the car".

Reinforce correct answer by saying,

Yes, these are just regular, literal sentences,

or correct wrong answer by indicating the right picture saying,

These are just regular, literal sentences.

Repeat with 'The cook is angry', providing the same feedback.

Continue with Novel Items. Do not provide feedback on response accuracy for test items.

HOOPER VISUAL ORGANIZATION TEST

This activity looks at your ability to recognise pictures of objects when the pictures have been cut up and placed in different positions. On each page you will see one cut up picture. Look at each picture and decide what it might be if it were put together. Then spell out the name of the object, using this alphabet board. Toward the end, the pictures become more difficult. Give an answer even if it is a guess.

MEMORY FOR FACES A

I am going to show you some pictures of faces, one at a time. Look at each face carefully and remember what it looks like.

Show faces for 2 seconds each. As you show each face, say

Remember this one.

Now I am going to show you some more pictures of faces, one at a time. I want you to look at the face on each page carefully. On this board, select Yes if the face is one that I asked you to remember or No if it is not.

Show faces. If participant says "I don't know", encourage them to guess. If they will not, record "DK" in the margin.

At end, say:

I want you to try to remember the first group of faces I asked you to remember because later on I'm going to ask you to pick them out of another group of faces.

LOGICAL MEMORY B

Do you remember the stories I read to you a little while ago? I am going to ask you some questions about the second story. For each question, select Yes or No from this card. If you are not sure of the answers, give your best guess.

GRADED ATTENTIONAL TEST

I am going to read you some lists of numbers. Each time, I will show you a card of what I want you to count in the list. For example, I might ask you to count how many times the number 5 appeared in the list. Then the card would look like this.

Show **Example Card A** [5].

Or, I might ask you to count all the times that 5 was followed immediately by a 3.

Show **Example Card B** [5 – 3].

Each time, all you have to do is answer with a single number, for the times the thing you were counting appeared. Give your answer by selecting it on this card once each list is finished.

[Show numbers response card, then cover during the reading of each list].

Any questions?

1) First, I want you to count the number of times 4 appears in this list.

Read list below, with [4] card displayed in book. Read numbers at one second intervals.

4 4 4 4 4 4 4 4

When finished, turn response card face up and say

How many times did the number 4 appear in the list?

Leave book at [4] card. For each subsequent trial, turn booklet to next page.

2) *Again, count the number of times 4 appears in this list.* Read list.

1 6 4 8 7 3 4 5 10 8 4 6 7 6 4

3) *This time, count the number of 4s and the number of 9s. Count them up as one single total, of the times that either number appears.*

3 3 1 6 3 8 9 6 2 4 7 2 7 8 9

4) *This time, count the number of times a 4, 9 or 6 is in the list. Again, count them up as a single total, of the times that any of these numbers appears.*

3 6 5 9 10 4 8 7 6 3 9 7 4 6 5

5) *This time, count the number of times there is a 4 followed immediately by a 9.*

4 8 7 4 9 3 7 4 9 7 3 9 4 8 3 4 9 5 4 5

6) *This time, count the number of times there is a 4 immediately followed by a 9, or a 6 followed immediately by a 2. Count it up as a single total, of the number of times either of these occurs.*

6 7 5 4 9 5 6 2 10 9 4 9 2 3 6 2 2 4 6 7

7) *Finally, count the number of times that there is a 4 followed by any digit, which is then immediately followed by a 9. For example, 4-1-9. Any time that there is a 4 followed by any number, followed immediately by a 9. Ready?*

7 4 3 10 2 4 5 9 2 4 6 9 3 1 6 3 10 4 3 5

AUDITORY RECEPTION

I'm going to read you a series of questions. For each one, you can answer either yes or no, by selecting your answer from this card. If you're not sure of any answer, just make your best guess.

For example, if I asked, "Do aeroplanes fly", *what would you answer?*

If correct, say

That's right, aeroplanes do fly.

If incorrect, say

Aeroplanes do fly, so the correct answer was yes. Let's try another one.

If I asked, "Do ponies shave", what would you answer?

If correct, say

That's right, ponies do not shave.

If incorrect, say

Ponies do not shave, so the correct answer was no.

Begin with test items.

MEMORY FOR FACES B

Now I'm going to show you some more pictures of faces. I want you to look at each face carefully. From this card, select Yes if the face is one that I asked you to remember earlier or No if it is not.

MATCH AND SHIFT CATEGORIES TEST

Trial 1 — Match by shape circle

I'm now going to show you a series of cards, and each card has four pictures on it. Your job to pick the picture that matches this card here

Indicate stimulus card.

Each time, I'll tell you whether you are right or wrong. If you're wrong, don't worry, we will just go on to the next card and try and get that one right.

OK, which picture matches this card here?

Stimulus Card 1 - One red circle. Matching by shape circle.

If correct, say

Yes, that matches in the way that I'm thinking of.

If incorrect, say

No, that doesn't match in the way that I'm thinking of.

Once participant gets **three** correct answers in a row, change response to just *Yes*, reverting to the longer answers if they again make an error (within the first 10 trials).

If participant fails to achieve three correct responses in a row in the first 10 cards, provide the following feedback after each incorrect response, from the 11th card:

No, that's not correct. This is the picture that matches in the way I'm thinking of.

Point to correct response.

Once 3 correct responses are made in a row revert to providing only *Correct/Incorrect* feedback. Continue to Card 30.

If by Card 20 participant has not made 3 correct responses in a row, discontinue trial and move on to Trial 2.

Trial 2 — Match by colour yellow

We're going to do the same thing now, but I want you to match the pictures to this card instead.

Turn book back to Card 1, display Stimulus 2 - Three yellow stars.

Show me which picture matches this card here.

If the participant successfully completed Trial 1, provide feedback as for Trial 1. Continue to Card 30, and then go to Trial 3.

If participant failed to achieve three correct responses in a row in Trial 1, provide the following feedback after each incorrect response, from the 1st card of Trial 2:

No, that's not correct. This is the picture that matches in the way I'm thinking of.

Point to correct response.

Once 3 correct responses are made in a row revert to providing only correct/incorrect feedback. If by Card 10 participant has not made 3 correct responses in a row on Trial 2, and did not complete Trial 1, discontinue test.

Trial 3 — Match by number two — Only begin if Trial 2 completed.

> *We're going to do the same thing once more, but I want you to match the pictures to a third card. Show me which picture matches this card.* Two blue triangles. Matching by number two.

Provide feedback on whether responses incorrect/correct. When incorrect do not indicate correct cards. Complete 30 cards regardless of performance.

Vary instructions as appropriate depending on administration order (i.e., Admin 3 vs. 4)

ED group: Break for the day if appropriate.

Else, disabled groups:

> *That's the end of the first part, which means we're half way though the activities we'll be doing today. Would you like to take a break at this point?*

Non-disabled groups:

> *That's the end of the first part, which means we're about a quarter/half* (depending on admin) *of the way through the activities. Would you like to take a break at this point?*

BLOCK B

FAMILY PICTURES A

This is a picture of the characters you will be seeing in the next <u>*two*</u> *scenes. I will be showing you* <u>*two*</u> *scenes with these family members and the dog in them, for 10 seconds. I want you to remember as much as you can about each scene. I will be asking you some questions about them later on.*

Identify the characters as follows and point to each character as you identify each. Say

This is the grandmother, grandfather, father, mother, son, daughter, and the dog.

Now I am going to show you the picnic scene and I want you to remember as much about the scene as you can.

Turn the page and expose for 10 seconds.

Now I am going to show you the department store scene. I want you to remember as much about the scene as you can.

Turn the page and expose for 10 seconds. DO NOT continue and show Scene 3.

Family names response card:

Who was in the picnic scene? <u>*Point to their names on this card.*</u>

Turn to the grid page after Scene 4. Point to the grid and say:

Pretend this is the picnic scene. You said the ________ [name the character identified by the participant] *was in the picnic scene. On this card, point to where that character was in the picture.*

Alphabet board

Now, <u>*using this alphabet board,*</u> *tell me what* ________ *was doing.*

Repeat for each character named.

Then ask,

Where there any other characters in the picnic scene?

If the participant says there were additional characters, repeat the questions above. Otherwise, continue with Scene 2, repeating the questions above, starting with:

Who was in the department store scene?

After completing the questioning for the two scenes, say:

Later on I will ask you questions about these scenes again, so try to remember them.

DIGIT SPAN FORWARDS

I am going to say some numbers. Listen carefully, and when I am through, I want you to <u>*repeat them back to me by using this number board*</u>*.*

Show participant number card, then cover card again.
Just repeat what I say.

For each trial, cover number card while numbers are being read, then uncover card to allow participant to make their response. Discontinue when errors are made on **both** trials of a particular span length (i.e., within the same shaded band on the record form).

DIGIT SPAN BACKWARDS

Now I am going to say some more numbers. But this time when I stop, I want you to repeat them backward. For example, if I say 2-9-5, what would you repeat back using the number card?

If participant responds correctly (5-9-2) say
That's right.
If the participant responds incorrectly, provide the correct response and say:
No, you would repeat 5-9-2. I said 2-9-5, so to repeat it backward, you would repeat 5-9-2. Now try these numbers. Remember, you are to repeat them backward: 1-6-3.

Whether or not the participant responds correctly (i.e., 3-6-1), proceed to Item 1. Discontinue when errors are made on **both** trials of a particular span length (i.e., within the same shaded band on the record form).

VERBAL PAIRED ASSOCIATES A

Response card: VPA Words
I am going to say a word, and then say another word that goes with it. I will say a whole list of words like that. Listen carefully because when I am finished I will say the first word, and I want you to tell me the word that goes with it, by pointing to one of the choices on this card.
Indicate words board, then hide board again.

Ensure the Words response card is not visible while the word lists are being read.
Now listen carefully to the list of word pairs as I read them.

If the participant beings to repeat the word pairs as you are reading, stop and instruct him or her to wait until you are finished. If the participant does not understand the directions, you may repeat them, paraphrasing where necessary.

Read the word pairs at the rate of one pair every 3 seconds; that is, the words are spoken about 1 second apart, with 2 seconds separating the pairs.

After reading List A, pause for 5 seconds and present Recall A. Show the Words response card, and ask:
Which word goes with _________ ?

If the participant responds correctly, say

Yes, _________ goes with _________

repeating the correct response. Then proceed to the next word in the Recall List.

If the participant responds incorrectly, say

No, _________ goes with _________

and provide the participant with the correct response.

The first time the participant makes an error, add

Each time, if you make a mistake, I'll tell you the correct answer so next time you're more likely to get it right.

Then proceed to the next word in the Recall List.

After completing List A, pause for 5 seconds. Then say

Now I will read the same list again, except with the word pairs in a different order. Listen carefully.

Read List B from the Record Form.

Use the same procedures for Lists B, C and D as for List A.

When the subtest is completed, say

Later on I will ask you to recall some of these word pairs again, so try to remember them.

FAMILY PICTURES B

Remember the pictures of family members doing things that I showed you earlier? I want you to remember which characters were in the scenes, where they were, and what they were doing.

Family names response card:

The first scene was the picnic scene. Who was in the picnic scene? Point to their names on this card.

Turn to the grid page in the WMS-III Stimulus Booklet 2. Point to the grid and say:

Pretend this is the picnic scene. You said the _________ [name the character identified] *was in the picnic scene. On this card, point to where that character was in the picture.*

Alphabet board

Now, using this alphabet board, tell me what _________ was doing.

Repeat for each character named.

Then ask,

Where there any other characters in the picnic scene?

If the participant says there were additional characters, repeat the questions above. Otherwise, continue with Scene 2, repeating the questions above, starting with:

Who was in the department store scene?

BOSTON NAMING TEST

> *I'm going to show you some pictures. For each picture, I'd like you to spell out on this alphabet board the name of the thing shown in the picture. It's not a spelling test, you just have to get across the general idea.*

Mark any reasonable communication of the exact word correct.

If participant gives an incorrect answer, or no answer, give stimulus clue, e.g.,

> *It's a musical instrument.*

If this does not help, give phonemic clue (part of word underlined on record sheet **only**) e.g.,

> *It starts with the sound "Acc".*

Complete whole subtest regardless of number of errors.

MATRIX REASONING

> *I am going to show you some pictures. For each picture, there is a part missing. Look at all aspects of each picture carefully and choose the missing part from the five choices.*

For Sample Items A-C, provide the teaching included in the Item Instructions if the participant responds incorrectly.

Place **Sample Item A** in front of the participant and say:

> *For example, tell me which of these pictures* (point to response choices) *should go here* (point to the question mark). *Make sure you carefully look at the picture on top and at the response choices below before making your selection. If you think there is more than one correct answer to the problem, choose the best one. Remember, you are to choose the one that best completes the pattern.*

If the participant responds correctly (Response 2), proceed to Sample Item B.
If the participant responds incorrectly, say:

> *For this item, the missing part should complete the pattern by making the picture the same colour. See, this choice* (Response 2) *would best complete the pattern because the squares are all yellow.*

Turn to **Sample Item B** and say:

> *Now tell me which of these pictures* (turn to the response choices) *should go here* (point to the question mark). *Again, make sure you carefully look at the picture on top and at the pictures below before choosing your answer. If you think that there is more than one correct answer to the problem, choose the best one.*

If the participant responds correctly (Response 5), proceed to Sample Item C.
If the participant responds incorrectly, say:

> *There are a number of ways you can solve this problem. For instance, you can look at the pictures by separating them into two columns. Notice the pictures in the left column are the same* (point at the two blue octagons).

> *They are both the same shape, and they are both blue. Now look at the right column* (point to the yellow circle and the question mark). *One of the choices below* (point to the response choices) *will make the pictures in the right column the same as well. See, this choice here* (Response 5) *would make the pictures in the right column both yellow circles.*

Turn to **Sample Item C** and say:

> *Now tell me which of these pictures* (point to the response choices) *should go here* (point to the question mark).

If the participant responds correctly (Response 4), proceed to Item 4.
If the participant responds incorrectly, say:

> *All the pictures at the top are circles, and each large circle is followed by a* (point to all the shapes in a sweeping motion) *small one. Therefore, the small circle* (Response 4) *is the best answer.*

Then proceed to Item 4 in the Stimulus Booklet.

Items 1-26

> *Now tell me which of these pictures* (point to response choices) *should go here* (point to the question mark).

No feedback or teaching should be given for Items 1-3 (if administered) and Items 4-26. The participant may make educated guesses, but instruct the participant not to guess randomly.

VERBAL PAIRED ASSOCIATES B

Response card: VPA Words

> *Remember the word pairs that you learned earlier? I told you a word and gave you another word that went with it. I want you to recall as many of those word pairs as you can remember, one more time. I will say the first word of the pair, and you tell me the word that goes with it, by pointing to one of the choices on this card.*

Indicate words board.

> *Which word goes with* _________ *?* (Read the first word of the pair).

DO NOT tell the participant if their response is correct or incorrect.

> *I am now going to read a list of word pairs. Listen carefully. After I read a word pair, I want you to select Yes from this card if it was one of the pairs that I asked you to remember earlier or NO if it is a new word pair. Some word pairs may be said more than once. Do you understand?*

Vary instructions as appropriate depending on administration order (i.e., Admin 3 vs. 4)

For non-disabled groups:

We're now about half way through the activities. Would you like a break?

or,

This might be a good point to stop today, and we'll organise another appointment to do the other half of the activities.

BLOCK C

BABCOCK STORY RECALL TEST A

I am going to read a short story to you now. Listen carefully because when I finish I'm going to ask you to tell me as much of the story as you can remember.

Read the story. Then,

Now tell me everything you can remember of the story.

Immediately following first recall trial, say,

In a little while I'm going to ask you to tell me how much of the story you can still remember. I'm going to read the story to you again now so that you'll have it fresh in your memory for the next time.

DIGIT-SYMBOL CODING

When we started today, I said that you will be doing all sorts of things. In this section, I'm going to ask you to copy some symbols.

Hand participant the record form.

Look at these boxes. Notice that each has a number in the upper part and a special mark in the lower part. Each number has its own mark.

Point to 1 and its mark in the key, then 2 and its mark. Then point to the seven squares located to the left of the heavy black line and say:

Now look down here where the squares have numbers in the top part but the squares at the bottom are empty. In each of the empty squares, put the mark that should go there. Like this.

Point to the first Sample Item, then point back to the key to show its corresponding mark, and say:

Here is a 2; the 2 has this mark. So I put it in the empty square, like this.

Write in the symbol. Point to the second Sample Item and say:

Here is a 1; the 1 has this mark (point to the second Sample Item, then to the mark below the 1 in the key), *so I put it in this square.*

Write in the symbol.

Point to the third sample item and say:

This number is a 3; the 3 has this mark (point to the third square and to the mark below the 3 in the key). *So I put it in the square* (write in the symbol).

After marking the first three Sample Items say:

Now you fill in the squares up to this heavy line.

If the participant makes an error on any of the Sample Items, correct the error immediately and review use of the key. When the participant completes a Sample Item correctly, offer encouragement by saying *Yes* or *Right.*

When all Sample Items have been completed, say:

> *Now you know how to do them. When I tell you to start, you do the rest of them.*

Point to the first square to the right of the heavy line and say:

> *Begin here and fill in as many squares as you can, one after the other without skipping any. Keep working until I tell you to stop. Work as quickly as you can without making any mistakes.*

Sweep across the first row with your finger and say:

> *When you finish this line, go on to this one.*

Point to the first square in the second row. Then point to the heavy black line and say:

> *Go ahead.*

Begin timing.
If the participant omits an item or starts to do only one type (e.g., only the 1's), say:

> *Do them in order. Don't skip any.*

At the end of 2 minutes (120 seconds), say *Stop.*

If the participant has not completed four rows at this point, mark their ending point (i.e., the item completed at 120 seconds) and then allow additional time for him or her to work to the end of the fourth row.

Record participant's ID number on top of sheet. If the participant completes the entire page in under 120 seconds, record the time to completion.

Incidental Learning
Remove Digit-Symbol Coding sheet, place Incidental Learning Sheet in front of participant.

> *Now I want you to fill in all the symbols you can remember that go with these numbers, one after another, across both rows. Tell me when you're finished.*

Use a blank sheet of paper to cover the Pairing items. Present the Free Recall task by pointing to the blank area at the bottom of the page. Say:

> *In this area* (point)*, I'd like you to write down all of the symbols you can remember, in any order. Tell me when you are finished.*

Symbol Copy
Record ID number on top of sheet, then remove Incidental Learning Sheet; place Symbol Copy sheet in front of participant.

> *This activity is similar to the one you did before, except this time simply copy the symbols into the square below. Work as quickly and as accurately as possible.*

At the end of 90 seconds, say *Stop.*

Record ID number on top of sheet. If the participant completes the entire page in under 90 seconds, record the time to completion.

HOOPER VISUAL ORGANIZATION TEST: COMPARISON MEASURE

> *As before, this activity looks at your ability to recognise pictures of objects when the pictures have been cut up and placed in different positions. On each page you will see one cut up picture. Look at each picture and tell me what it might be if it were put together. Toward the end, the pictures become more difficult. Give an answer even if it is a guess.*

FINGER TAPPING TEST

> *This is a activity to see how fast you can move your fingers. You will have a number of attempts with each hand. When you are moving the tapper, like this* (demonstrate), *you must ensure that it goes all the way to the top, and all the way to the bottom, in order to make it click over. Try to go as fast as you can. Let's start with your* [dominant] *hand. Start when I say go, and keep going till I stay stop. Ready? Go!*

Alternate trials on each hand; total of five trials of 10 seconds for each hand. Ensure participant is using their index finger, and is only moving the finger while the rest of the hand remains on the table. Do not allow participant to move their entire hand from the wrist.

BABCOCK STORY RECALL TEST B

> *You recall the story I read you a little while ago. Tell me everything you can remember from the story.*

WISCONSIN CARD SORTING TEST

> *This test is a little unusual because I am not allowed to tell you very much about how to do it. You will be asked to match each of the cards in these decks* (point to the response card decks) *to one of these four key cards* (point to each of the stimulus cards in succession, beginning with the red triangle). *You must always take the top card from the deck and place it below the key card you think it matches. I cannot tell you how to match the cards, but I will tell you whether you are right or wrong. If you are wrong, simply leave the card where you have placed it and try to get the next card correct. There is no time limit on this test. Are you ready? Let's begin.*

Vary instructions as appropriate depending on administration order (i.e., Admin 3 vs. 4)

> *We're now three-quarters of the way through the activities. Would you like to take a break at this point?*

BLOCK D

FAMILY PICTURES A: COMPARISON MEASURE

> *This is a picture of the characters you will again be seeing in two scenes. I will be showing you two scenes with these family members and the dog in them, for 10 seconds. I want you to remember as much as you can about each scene. I will be asking you some questions about them later on.*

Identify the characters as follows and point to each character as you identify each. Say

> *Remember, this is the grandmother, grandfather, father, mother, son, daughter, and the dog.*

> *Now I am going to show you the yard scene and I want you to remember as much about the scene as you can.*

Turn the page and expose for 10 seconds.

> *Now I am going to show you the meal scene. I want you to remember as much about the scene as you can.*

Turn the page and expose for 10 seconds.

> *Who was in the yard scene?*

Turn to the grid page after Scene 4. Point to the grid and say:

> *Pretend this is the yard scene. You said the* ________ [name the character identified by the participant] *was in the yard scene. On this card, point to where that character was in the picture.*

> *Now, tell me what* ________ *was doing.*

Repeat for each character named.

Then ask,

> *Where there any other characters in the yard scene?*

If the participant says there were additional characters, repeat the questions above. Otherwise, continue with Scene 2, repeating the questions above, starting with:

> *Who was in the meal scene?*

After completing the questioning for the two scenes, say:

> *Later on I will ask you questions about these scenes again, so try to remember them.*

DIGIT SPAN FORWARDS: COMPARISON MEASURE

> *I am going to say some numbers. Listen carefully, and when I am through, I want you to say them right after me. Just say what I say.*

Discontinue when errors are made on **both** trials of a particular span length (i.e., within the same shaded band on the record form).

Digit Span Backwards: Comparison Measure

Now I am going to say some more numbers. But this time when I stop, I want you to say them backward. For example, if I say 7-1-9, what would you say?

If participant responds correctly (9-1-7) say *That's right.*

If the participant responds incorrectly, provide the correct response and say:

No, you would say 9-1-7. I said 7-1-9, so to say it backward, you would say 9-1-7. Now try these numbers. Remember, you are to say them backward: 3-4-8.

Whether or not the participant responds correctly (i.e., 8-4-3), proceed to Item 1.

Discontinue when errors are made on **both** trials of a particular span length (i.e., within the same shaded band on the record form).

Auditory Verbal Learning Test A

I am going to read a list of words. Listen carefully, for when I stop you are to repeat back as many words as you can remember. It doesn't matter in what order you repeat them. Just try to remember as many as you can.

Read List A words, with a 1-second interval between each of the 15 words. Check of the words recalled using numbers to keep track of the pattern of recall. No feedback should be given regarding the number of correct responses, repetitions, or errors.

When the participant indicates that he or she can recall no more words, re-read the list following this second set of instructions:

Now I am going to read the same words again, and once again when I stop I want you to tell me as many words as you can remember, including words you said the first time. It doesn't matter in what order you say them. Just say as many words as you can remember whether or not you said them before.

The list is reread for Trials 3 through 5 using Trial 2 instructions each time. After Trial 5, read List B with the following instructions:

Now I'm going to read a second list of words. Listen carefully, for when I stop you are to repeat back as many words as you can remember. It doesn't matter in what order you repeat them. Just try to remember as many as you can.

Immediately after the List B trial, ask the patient to recall as many words from the first list (List A) without further presentation of those words:

Now tell me all the words that you can remember from the first list.

Family Pictures B: Comparison Measure

Remember the second two pictures of family members doing things that I showed you recently? I want you to remember which characters were in the scenes, where they were, and what they were doing.

Who was in the yard scene?

Turn to the grid page in the WMS-III Stimulus Booklet 2. Point to the grid and say:
Pretend this is the yard scene. You said the ________ [name the character identified] *was in the yard scene. On this card, point to where that character was in the picture.*

Now, tell me what ________ *was doing.*

Repeat for each character named.

Then ask, *Where there any other characters in the yard scene?*
If the participant says there were additional characters, repeat the questions above. Otherwise, continue with Scene 4, repeating the questions above, starting with:
Who was in the meal scene?

Boston Naming Test: Comparison Measure

I'm going to show you some pictures. For each picture, I'd like you to tell me the name of the thing shown in the picture.

Block Design
(Start with Design 5)

Reverse Basal Items Instructions
Trial 1 (Time Limit: 30")
With four blocks ready, say:
I am going to put these blocks together and make a design. Watch me.

Now make one just like this (point to the model). *Tell me when you are finished. Go ahead.*

Trial 2 (Time Limit: 30")
Leave your model intact. Using the participant's two blocks to demonstrate and explain the construction of the model, say:
Watch me again.

Then scramble the participant's blocks and say:
Now, you try it again and be sure to make it just like mine.

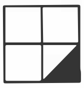
Begin timing. Whether the participant succeeds or fails on the second trial say:
Let's try a new one.

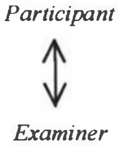
Design 1

Design 2

Design 3

Design 4





START HERE

Design 5

Trial 1 (Time Limit: 60")

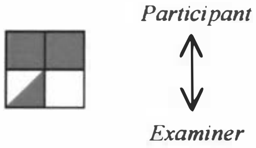
Place four blocks in front of the participant and say:

> *Now I'm going to ask you to make some designs. You see these blocks? They are all alike. On some sides they are all red; on some, all white; and on some, half red and half white.*

Turn the blocks to show the different sides. Take out four more blocks. Say:

> *I am going to put some blocks together to make a design. Watch me.*

Construct the model of Design 5 using the four blocks:



> *Now make one just like this. Tell me when you've finished.*

Trial 2 (ONLY IF FAILED ON TRIAL 1; Time Limit: 60")

> *Watch me again.*

Leaving your model intact and using the participant's blocks, demonstrate the correct construction of the design for a second time. Then, scramble the participant's blocks and say:

> *Now you try it again and be sure to make it just like mine* (point to the model). *Tell me again when you are finished.*
>
> Begin timing. Proceed to Design 4 regardless of performance on Trial 2 of Design 5. Administer Designs 1-4 in reverse sequence until the participant obtains perfect scores (2 points) on two consecutive items. When this criterion is met, give full credit for any proceeding items that were not administered. Then proceed with Trial 1 of Design 6.

Design 6

Trial 1 (Time Limit: 60")

Scramble four blocks in front of the participant.

Turn to Design 6 in the Stimulus Booklet and place it in front of the participant. Be sure the unbound edge is closest to the participant. Say:

> *This time we are going to put the blocks together to make them look like this picture* (point to the design shown in the Stimulus Booklet). *Watch me first.*

Slowly demonstrate the construction of the design using the participant's blocks. Upon completion, say:

> *You see, the tops of the blocks look the same as this picture.*

Scramble the blocks used in the demonstration and say:

> *Now look at the picture and make one just like it with these blocks. Tell me when you are finished. Go ahead.*

Begin timing. If the participant successfully constructs the design within the time limit, proceed to Design 7. If the participant constructs the design incorrectly, proceed to Trial 2.

Trial 2 (ONLY IF FAILED ON TRIAL 1; Time Limit: 60")

Watch me again.

Construct a model of Design 6, using the participant's blocks. Then, scramble the blocks and say:

Now you try it again and be sure to make it just like this one on this page.

Begin timing. Then, proceed to Design 4 regardless of performance on Trial 2 of Design 6. Administer Designs 1-4 in reverse sequence until the participant obtains perfect scores (2 points) on two consecutive items. When this criterion is met, give full credit for any proceeding items that were not administered. Then proceed with the subtest until the discontinue criterion is met.

Designs 7-9 (Time Limit: 60" Each Design)

Scramble four blocks in front of the participant.

Turn the Stimulus Booklet page to expose the next design and say:

Now make one just like this. Try to work as quickly as you can. Tell me when you are finished.

Designs 10-14 (Time Limit: 120" Each Design)

Scramble nine blocks in front of the participant.

Turn the Stimulus Booklet page to expose the next design and say:

Now make one just like this using nine blocks. Be sure to tell me when you have finished.

AUDITORY VERBAL LEARNING TEST B

A while ago, I read a list of words to you several times, and you had to repeat back the words. Tell me the words from that first list.

After list is recalled, show participant recognition response card.

This list shows a whole lot of words, including the words from the first list, the one which I read five times. Tell me all the words on this page that are from that first list.

[Participant looks at response card, but verbally names words which they recognise].

After participant has finished naming words, say

Also in this list are the words from the second list I read you, the one I read only once. Tell me all the words that are from that second list.

WRAP UP

Thank you very much for participating in this research. This research is important and we value your contribution.

We have $15 for you, in partial recognition for your effort and expenses participating in the research. Here is the $15 (put money on the table). *I need you to sign this form saying that you have received the $15 from me.*

If in the follow-up group (first half of participants)

Also, we are asking participants if they would be willing to participate in the second stage of the research. The second stage would involve a another assessment, similar to the one today, although it would be only about half as long. Even though it is half as long, you will still be paid $15 for your expenses for the second assessment as well. If you are going to participate in the second part it would need to be conducted in about 2 weeks time. You are under no obligation to participate in this second part.

Would you be interested in being involved in the second part?

If yes, make appointment, ideally for 14 -21 days time. Do not book closer than 14 days away, or further than 28 days away.

Also, if you wish we can provide you with a short summary of the results of the research if you would like. This would be sent out towards the end of this year. If you would like a summary sent to you, you need to tick this box (indicate box) *and write your name and address here* (point to space at bottom of page). *This information is confidential, and will be kept separate from your test results.*

Thanks very much. [See you in two/three weeks].

APPENDIX D. NORMATIVE GROUP SCREENING PROCEDURE

1. *Describe nature of study, time involvement required, and money for expenses.*

2. Would you be potentially interested in participating in the research?

3. There are a number of specific criteria for participating in the research. Perhaps I could just run through those with you now…

   If meets any of the following criteria, probe for further detail. If exclusionary criteria met, go to 5 below:
   Have you ever had a brain injury? *(If yes, check severity).*
   Have you ever been knocked unconscious? *(Exclude if LOC > 5 minutes)*
   Have you had a stroke?
   Have you had a heart attack?
   Have you been diagnosed with a brain tumour?
   Have you been diagnosed with epilepsy?
   Many people don't know this, but do you know if you had an episode at your birth where you turned blue or had limited oxygen? *(Only exclude if yes, not if DK)*
   Have you ever been hospitalised for a psychiatric condition?

   If meets any of the following, probe for further information, and consider whether to exclude. If yes, go to 5 below.
   Have you had a concussion without unconsciousness?
   Have you ever required a CT scan or MRI of your head?
   Are you taking any regular medications, (*if female:* other than contraceptives)?
   Do you use any non-prescription drugs, such as cannabis?

   What is your age, please?
   Record age. If age less than 16 or greater than 55, go to 5 below.

   What is your highest level of formal education?
   Record education.

4. OK. You would be suitable to participate in the research. Can we make a time for you to be seen?
   Assign ID number and administration, and arrange an appointment time for participant to be seen.

If participant is unsuitable, say:

5. I'm sorry, but you don't meet all of the criteria for this research. Thank you for your interest in this project.

APPENDIX E. RESEARCH PROTOCOL DATASHEET

This datasheet for Administration 3 is provided as an example, and was one of four administration formats (see *Counterbalancing* in Chapter 7, *Methodology*). This datasheet is presented at 80% of its original size.

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

NEUROPSYCHOLOGICAL ASSESSMENT OF INDIVIDUALS WITH EXPRESSIVE DISABILITIES IN ADDITION TO TRAUMATIC BRAIN INJURIES

Research Protocol Data Sheet

ADMN 3

DEMOGRAPHICS A

ID Number ______________________ **ID**

Date ______________________ **DATE**

Location ______________________ Location code: ________ **LOC**

Assessor Surname ______________________ Assessor code: ________ **ASSOR**

Group **GROUP**

	TBI	No TBI
Disability	1	2
No Disability	3	4

Response mode Pointing 1 Eye movements 2 **RESP**

Sex Female 1 Male 2 **SEX**

Date of Accident ______________________ **DTA**

Cause of Accident ______________________ Cause code: ________ **COA**

Assessment Initial 1 Follow-up 2 **ASS**

Manual Time Calculation Required No 1 Yes 2 **MANT**

INFORMATION SHEET AND CONSENT FORMS

INFT ________ 8

Te Kunenga ki Pūrehuroa
Inception to Infinity: Massey University's commitment to learning as a life-long journey

RESPONSE RELIABILITY (CARDS 1–7)

RRT _______ ⧖

Yes/No Card **RR1-12**

1.	No	0	1
2.	I Don't Know	0	1
3.	Yes	0	1
4.	I Made A Mistake	0	1

Numbers Card

5.	4	0	1
6.	6	0	1
7.	3	0	1
8.	7	0	1

Family Card

9.	Grandfather	0	1
10.	Father	0	1
11.	Grandmother	0	1
12.	Daughter	0	1

Words Card **RR13-20**

13.	Acorn	0	1
14.	Ladder	0	1
15.	Clown	0	1
16.	Glass	0	1

Picture Numbers Card

17.	4	0	1
18.	3	0	1
19.	2	0	1
20.	1	0	1

Colour Card **RR21-28**

21.	Blue	0	1
22.	Green	0	1
23.	Yellow	0	1
24.	Red	0	1

Picture Form Card

25.	Circle	0	1
26.	Square	0	1
27.	Star	0	1
28.	Triangle	0	1

DEMOGRAPHICS B (ETHNICITY – CARD 8)

DEMT _______ ⧖

Ethnicity	Pakeha 1 Māori 2 Pacific Island 3 Asian 4 Other European 5		**ETH**
Age	______________________________		**AGE**
Employment	______________________________	Employment type: ________	**EMP**
Education	______________________________	Education years: ________	**EDUC**
Handedness	Left 1 Right 2 N/A 3		**LAT**
Colour-blind	No 1 Yes 2		**COL**

MEDICAL SCREEN (YES/NO – CARD 1)

MEDT _______ ⧖

		MED1-12 **N**	 **Y**
1.	Head injury	0	1
2.	Knocked unconscious	0	1
3.	Stroke	0	1
4.	Heart attack	0	1
5.	Diagnosed with a brain tumour	0	1
6.	Diagnosed with epilepsy	0	1
7.	Episode at birth where turned blue/had limited oxygen	0	1
8.	Have you ever been hospitalised for a psychiatric condition?	0	1
9.	Concussion without unconsciousness (specify number, and duration of longest state of confusion in minutes: ______ ____)	0	1
10.	Required a CT scan or MRI of your head? (specify reason: ________________________)	0	1

11.	Taking any regular medications, (excl. contraceptives), prescribed or self-administered?	0	1
12.	Use any non-prescription drugs (such as cannabis)?	0	1

If the answer to any of the above is "Yes", then examine sample exclusionary criteria and consider whether participant should be excluded from the sample.

GOOD SAMARITAN HOSPITAL ORIENTATION TEST (YES/NO – CARD 1)

ORT _______ ⧗

OR1–10

		0	**1**
Personal Information			
1.	Is your name Jo(e)? (incorrect; if Jo or variant is real name then Sam)	Y	N
2.	Is your name… ? (correct first name; do not record here)	N	Y
3.	Are you _______ years old? (correct)	N	Y
4.	Are you _______ years old? (incorrect, actual +5 years)	Y	N
5.	Do you live in ________________? (correct region)	N	Y
6.	Do you live in Canterbury?	Y	N
Orientation to Place			
7.	Is this place a (incorrect; circle one of: hospital, school, clinic, community home)?	Y	N
8.	Is this place a ________________? (correct; describe place)	N	Y
Orientation to Time			
9.	Is it the month of ________________ now? (actual -3 mths, *see list below*) *[Jan=Oct, Feb=Nov, Mar=Dec, Apr=Jan, May=Feb, Jun=Mar, Jul=Apr, Aug=May, Sep=Jun, Oct=Jul, Nov=Aug, Dec=Sep]*.	Y	N
10.	Is it the month of ________________ now? (correct)	N	Y

BLOCK A

LOGICAL MEMORY A (YES/NO – CARD 1)

LMAT ________ 8

		LMA1-15	
		0	**1**
1.	Was the woman's name Diana Thompson?	Y	N
2.	Was the story setting in South Boston?	N	Y
3.	Was the woman a cook?	N	Y
4.	Did she work in a restaurant?	Y	N
5.	Did she have four children?	N	Y
6.	Were the children teenagers?	Y	N
7.	Did the robbery take place on Sixth Street?	Y	N
8.	Did the woman report being robbed two nights before?	Y	N
9.	Did she report the robbery at the Police Station?	N	Y
10.	Was the woman robbed of 75 dollars?	Y	N
11.	Did the family go without food for four days?	Y	N
12.	Was the rent due?	N	Y
13.	Did the police catch the thief?	Y	N
14.	Did the police feel sorry for the woman?	N	Y
15.	Did the police take up a collection?	N	Y

READ SECOND STORY! ☐

FAMILIAR AND NOVEL LANGUAGE COMPREHENSION TEST

FNLT ________ 8

Scoring Key:

1	2
3	4

A.	He's got his head in the clouds. [2]	**FNLN1-20**	
B.	The coast is clear. [3]	**0**	**1**
1.	I'd like to give you a piece of my mind.	Other	3
2.	Sticks and stones will break my bones but words will never hurt me.	Other	1
3.	She took a sudden turn for the worse.	Other	1
4.	Just in the nick of time.	Other	2
5.	It's like talking to a brick wall.	Other	1
6.	I've got a bone to pick with you.	Other	4
7.	I'll get back to you later.	Other	1
8.	The truth, the whole truth, and nothing but the truth.	Other	4
9.	He's saving up for a rainy day.	Other	3
10.	When our ship comes in.	Other	4
11.	She's got him eating out of her hand.	Other	4
12.	That's enough to drive a man to drink.	Other	1
13.	While the cat's away, the mice will play.	Other	2
14.	Keep your nose to the grindstone.	Other	2
15.	It seems like just yesterday.	Other	3
16.	Why don't you pick on someone your own size?	Other	4
17.	He's living high on the hog.	Other	3
18.	Cat got your tongue?	Other	3
19.	How about a little peace and quiet around here?	Other	4
20.	He's turning over a new leaf.	Other	2

Scoring Key:

1	2
3	4

A.	He takes his pets in the car. [1]	**FNLF1-20**	
B.	The cook is angry. [3]	0	1
1.	The dog's trying to give her a ride on the wagon.	Other	4
2.	The nails are under the square and the hammer is in the circle.	Other	3
3.	He's racing a truck against a horse.	Other	1
4.	Almost to the bottom of the mountain.	Other	3
5.	She tried jumping over the striped cat.	Other	1
6.	He's got a picture to show of her.	Other	4
7.	He jumped up to her suddenly.	Other	2
8.	The clown, the small clown, and not the one near a girl.	Other	3
9.	She's looking down at her black cat.	Other	1
10.	Then her dog walks in.	Other	4
11.	He sees her drinking from a bowl.	Other	4
12.	It's easy to teach a dog to swim.	Other	3
13.	Whenever the sun sets, the dog barks.	Other	2
14.	Follow your sister to the dinner table.	Other	4
15.	He kisses the thin lady.	Other	3
16.	Where are they not showing each other their own hats?	Other	2
17.	He's sitting deep in the bubbles.	Other	2
18.	Who's following the dog?	Other	1
19.	Show me the apple and large banana above a square.	Other	2
20.	He's chasing after a white duck.	Other	4

HOOPER VISUAL ORGANIZATION TEST (ALPHABET – CARD 9)

VOT ______ ⌛

	0-point response	**1-point response**	**2-point response**	**VO1-15**		
1.			Fish	0	1	2
2.			Table, Bench	0	1	2
3.		Football	Baseball, other round ball	0	1	2
4.		Animal	Dog, Sheep	0	1	2
5.		Vase, Jar	Cup	0	1	2
6.		Fruit	Apple, Peach, etc.	0	1	2
7.			Scissors	0	1	2
8.			Sailboat, Boat	0	1	2
9.		Sofa	Chair	0	1	2
10.			Teapot, Cream Pitcher	0	1	2
11.			Flower, Pansy, etc.	0	1	2
12.			Book	0	1	2
13.			Block	0	1	2
14.	Iron		Shoe	0	1	2
15.	Lock		Ring	0	1	2

MEMORY FOR FACES A (YES/NO – CARD 1)

MFAT _______ ⧗

	MFA1-12			MFA13-25			MFA25-36			MFA37-48	
	0	1		0	1		0	1		0	1
1.	Y	N	13.	Y	N	25.	N	Y	37.	N	Y
2.	N	Y	14.	N	Y	26.	Y	N	38.	N	Y
3.	Y	N	15.	Y	N	27.	Y	N	39.	Y	N
4.	Y	N	16.	N	Y	28.	N	Y	40.	N	Y
5.	N	Y	17.	Y	N	29.	N	Y	41.	Y	N
6.	Y	N	18.	Y	N	30.	Y	N	42.	N	Y
7.	N	Y	19.	N	Y	31.	Y	N	43.	Y	N
8.	Y	N	20.	Y	N	32.	N	Y	44.	N	Y
9.	Y	N	21.	N	Y	33.	N	Y	45.	Y	N
10.	Y	N	22.	N	Y	34.	Y	N	46.	N	Y
11.	N	Y	23.	Y	N	35.	N	Y	47.	N	Y
12.	N	Y	24.	N	Y	36.	Y	N	48.	Y	N

LOGICAL MEMORY B (YES/NO – CARD 1)

LMBT _______ ⧗

		LMB1-15	
		0	1
1.	Was the man's name Joe Garcia?	N	Y
2.	Was it Sunday evening?	Y	N
3.	Was it 6:00?	N	Y
4.	Was the story setting in Seattle?	Y	N
5.	Was Joe dressing to go out?	N	Y
6.	Was Joe watching television?	N	Y
7.	Was the program interrupted?	N	Y
8.	Was the storm expected to move into the area on Tuesday?	Y	N
9.	Was the storm expected to stay in the area through the night?	N	Y
10.	Was the temperature predicted to drop 30 degrees?	Y	N
11.	Did the announcer predict 10 inches of rain?	Y	N
12.	Did the announcer warn of possible flooding?	Y	N
13.	Did the announcer warn that it could hail?	N	Y
14.	Did Joe decide to stay home?	N	Y
15.	Did Joe sit down to watch a sports program?	Y	N

GRADED ATTENTIONAL TEST (NUMBERS – CARD 2)

Discontinue:
Three scores of 0 in a row.

GAT _______ ⧗

		GA1-4	
		0	1
1.	4s in sequence of 4s	Other	8
2.	4s in mixed sequence	Other	4
3.	4s and 9s	Other	3
4.	4s, 9s and 6s	Other	7

		GA5-7	
		0	1
5.	Sequence pairs of 4–9	Other	3
6.	Pairs of 4–9 and 6–2	Other	4
7.	Sequences of 4–*–9	Other	2

AUDITORY RECEPTION (YES/NO – CARD 1)

Reverse Basal Items (if required)

		AR1-10	
		0	1
1.	Do dogs eat?	N	Y
2.	Do dogs fly?	Y	N
3.	Do trees fly?	Y	N
4.	Do babies drink?	N	Y
5.	Do babies cry?	N	Y
6.	Do bicycles eat?	Y	N
7.	Do dresses sing?	Y	N
8.	Do children climb?	N	Y
9.	Do cats bark?	Y	N
10.	Do bees sting?	N	Y

Demo

a.	Do aeroplanes fly?	N	Y
b.	Do ponies shave?	Y	N

Test Items

		AR11-22	
		0	1
11.	Do people marry?	N	Y
12.	Do bananas telephone?	Y	N
13.	Do ants crawl?	N	Y
14.	Do eagles paint?	Y	N
15.	Do bricks float?	Y	N
16.	Do hatchets chop?	N	Y
17.	Do dials yawn?	Y	N
18.	Do logs burn?	N	Y
19.	Do sidewalks sprinkle?	Y	N
20.	Do penguins waddle?	N	Y
21.	Do pincushions cheer?	Y	N
22.	Do sausages frown?	Y	N

Reverse:
If score of 0 on any of Items 11-15, administer Items 1-10 in reverse until 5 consecutive correct answers.

ART ________ ⌛

		AR23-50	
		0	1
23.	Do flowers bloom?	N	Y
24.	Do parachutes paddle?	Y	N
25.	Do scouts signal?	N	Y
26.	Do clowns tumble?	N	Y
27.	Do bugles camouflage?	Y	N
28.	Do chimneys relax?	Y	N
29.	Do magicians entertain?	N	Y
30.	Do barometers congratulate?	Y	N
31.	Do dentists drill?	N	Y
32.	Do microscopes magnify?	N	Y
33.	Do zebras burrow?	Y	N
34.	Do brides dream?	N	Y
35.	Do cosmetics celebrate?	Y	N
36.	Do leaves flutter?	N	Y
37.	Do portals precipitate?	Y	N
38.	Do carpenters kneel?	N	Y
39.	Do sphinxes gallop?	Y	N
40.	Do meteorites collide?	N	Y
41.	Do dictionaries define?	N	Y
42.	Do weasels knit?	Y	N
43.	Do combustibles ignite?	N	Y
44.	Do canines manufacture?	Y	N
45.	Do beverages quench?	N	Y
46.	Do pigeons coo?	N	Y
47.	Do efficient compasses misdirect?	Y	N
48.	Do wingless birds soar?	Y	N
49.	Do migratory birds traverse?	N	Y
50.	Do mute musicians vocalise?	Y	N

MEMORY FOR FACES B (YES/NO – CARD 1)

MFBT ________ ⌛

	MFB1-12	
	0	1
1.	Y	N
2.	N	Y
3.	Y	N
4.	Y	N
5.	N	Y
6.	Y	N
7.	N	Y
8.	Y	N
9.	Y	N
10.	Y	N
11.	N	Y
12.	Y	N

	MFB13-24	
	0	1
13.	N	Y
14.	Y	N
15.	N	Y
16.	Y	N
17.	Y	N
18.	N	Y
19.	Y	N
20.	N	Y
21.	N	Y
22.	Y	N
23.	N	Y
24.	Y	N

	MFB25-36	
	0	1
25.	N	Y
26.	Y	N
27.	N	Y
28.	N	Y
29.	N	Y
30.	Y	N
31.	N	Y
32.	Y	N
33.	N	Y
34.	Y	N
35.	N	Y
36.	N	Y

	MFB37-48	
	0	1
37.	Y	N
38.	Y	N
39.	N	Y
40.	N	Y
41.	Y	N
42.	N	Y
43.	N	Y
44.	Y	N
45.	Y	N
46.	N	Y
47.	Y	N
48.	N	Y

MATCH AND SHIFT CATEGORIES TEST (CARDS 10–12)

MST _______ ⧗

	MSF1-15	MSC1-15	MSN1-15		MSF16-30	MSC16-30	MSN16-30
	1 2 3 4	1 2 3 4	1 2 3 4		1 2 3 4	1 2 3 4	1 2 3 4
1.	A B **C** D	**A** B C D	A B C **D**	16.	A **B** C D	**A** B C D	A B C **D**
2.	A B **C** D	A B C **D**	**A** B C D	17.	**A** B C D	A B C **D**	A **B** C D
3.	**A** B C D	**A** B C D	**A** B C D	18.	A **B** C D	A B C **D**	A **B** C D
4.	A B **C** D	A B **C** D	A B C **D**	19.	**A** B C D	**A** B C D	A B C **D**
5.	A **B** C D	A B **C** D	A B C **D**	20.	A B **C** D	**A** B C D	A B **C** D
6.	A B **C** D	A **B** C D	A B C **D**	21.	A **B** C D	A B C **D**	**A** B C D
7.	A B **C** D	A **B** C D	**A** B C D	22.	A B C **D**	A B **C** D	A **B** C D
8.	A B C **D**	A B **C** D	**A** B C D	23.	A **B** C D	A **B** C D	A B **C** D
9.	A **B** C D	A B **C** D	A B **C** D	24.	**A** B C D	A B C **D**	A B C **D**
10.	A B **C** D	A B **C** D	**A** B C D	25.	**A** B C D	A **B** C D	**A** B C D
11.	**A** B C D	**A** B C D	**A** B C D	26.	A B C **D**	**A** B C D	A B **C** D
12.	A **B** C D	A B **C** D	A **B** C D	27.	A B **C** D	A **B** C D	**A** B C D
13.	A **B** C D	A B C **D**	A **B** C D	28.	A B **C** D	A **B** C D	**A** B C D
14.	A **B** C D	**A** B C D	A B **C** D	29.	A B **C** D	A B C **D**	A B **C** D
15.	A **B** C D	A **B** C D	A **B** C D	30.	A **B** C D	A B C **D**	A **B** C D

BKAT _______ ⧗

BLOCK B

FAMILY PICTURES A (FAMILY NAMES – CARD 3; ALPHABET – CARD 9)

FPAT _______ ⧗

Scene A **Picnic**		**Present** **FPA1-7**			**Location** **FPA8-14**			**Activity** **FPA15-21**				
		0	1		0	1		0	1	2		
Grandmother	1.	Y	N	8.	1 2 3 4	-	15.	0	-	-		-
Grandfather	2.	N	Y	9.	1 2 4	3	16.	0	1	2		Watching dog catch frisbee
Mother	3.	Y	N	10.	1 2 3 4	-	17.	0	-	-		-
Father	4.	N	Y	11.	2 3 4	1	18.	0	1	2		BBQing hamburgers
Daughter	5.	N	Y	12.	1 3 4	2	19.	0	1	2		Throwing frisbee
Son	6.	Y	N	13.	1 2 3 4	-	20.	0	-	-		-
Dog	7.	N	Y	14.	1 2 3	4	21.	0	1	2		Catching frisbee

Scene B **Dept. Store**		**Present** **FPA22-28**			**Location** **FPA29-35**			**Activity** **FPA36-42**				
		0	1		0	1		0	1	2		
Grandmother	22.	Y	N	29.	1 2 3 4	-	36.	0	-	-		-
Grandfather	23.	N	Y	30.	1 2 3	4	37.	0	1	2		Pulling out sweater
Mother	24.	N	Y	31.	1 3 4	2	38.	0	1	2		Looking in purse
Father	25.	N	Y	32.	1 2 4	3	39.	0	1	2		Looking at shirt
Daughter	26.	N	Y	33.	2 3 4	1	40.	0	1	2		Looking in mirror
Son	27.	Y	N	34.	1 2 3 4	-	41.	0	-	-		-
Dog	28.	Y	N	35.	1 2 3 4	-	42.	0	-	-		-

DIGIT SPAN FORWARDS (NUMBERS – CARD 2)

DSFT _______ ⧗

		DSF1-10	
1.	7 – 4	0	1
2.	5 – 7	0	1
3.	8 – 7 – 1	0	1
4.	3 – 6 – 7	0	1
5.	3 – 4 – 8 – 7	0	1
6.	3 – 6 – 4 – 8	0	1
7.	8 – 6 – 9 – 3 – 1	0	1
8.	3 – 8 – 4 – 5 – 2	0	1
9.	7 – 1 – 5 – 8 – 2 – 3	0	1
10.	6 – 4 – 7 – 8 – 5 – 3	0	1

		DSF11-16	
11.	7 – 1 – 6 – 3 – 9 – 4 – 5	0	1
12.	5 – 7 – 3 – 8 – 9 – 1 – 2	0	1
13.	4 – 3 – 7 – 2 – 6 – 5 – 8 – 9	0	1
14.	6 – 5 – 8 – 2 – 7 – 1 – 4 – 3	0	1
15.	3 – 7 – 6 – 8 – 5 – 2 – 9 – 4 – 1	0	1
16.	2 – 4 – 5 – 8 – 9 – 3 – 1 – 7 – 6	0	1

DIGIT SPAN BACKWARDS (NUMBERS – CARD 2)

DSBT ______ 8

			DSB1-14	
1.	7 – 9	(9 – 7)	0	1
2.	5 – 4	(4 – 5)	0	1
3.	3 – 7 – 4	(4 – 7 – 3)	0	1
4.	5 – 9 – 4	(4 – 9 – 5)	0	1
5.	4 – 8 – 6 – 1	(1 – 6 – 8 – 4)	0	1
6.	2 – 1 – 6 – 9	(9 – 6 – 1 – 2)	0	1
7.	6 – 7 – 1 – 3 – 4	(4 – 3 – 1 – 7 – 6)	0	1
8.	1 – 8 – 3 – 2 – 9	(9 – 2 – 3 – 8 – 1)	0	1
9.	6 – 7 – 4 – 9 – 3 – 2	(2 – 3 – 9 – 4 – 7 – 6)	0	1
10.	7 – 8 – 5 – 2 – 3 – 4	(4 – 3 – 2 – 5 – 8 – 7)	0	1
11.	1 – 5 – 2 – 3 – 7 – 4 – 8	(8 – 4 – 7 – 3 – 2 – 5 – 1)	0	1
12.	1 – 8 – 3 – 7 – 2 – 6 – 5	(5 – 6 – 2 – 7 – 3 – 8 – 1)	0	1
13.	5 – 9 – 8 – 2 – 6 – 4 – 3 – 7	(7 – 3 – 4 – 6 – 2 – 8 – 9 – 5)	0	1
14.	5 – 7 – 3 – 9 – 4 – 2 – 8 – 6	(6 – 8 – 2 – 4 – 9 – 3 – 7 – 5)	0	1

VERBAL PAIRED ASSOCIATES A (OBJECT WORDS – CARD 4)

VPAT ______ 8

List A	Recall A	VPAA1-8	
Truck–Arrow	1. Bank (Cartoon)	0	1
Insect–Acorn	2. Reptile (Clown)	0	1
Reptile–Clown	3. Star (Ladder)	0	1
Bank–Cartoon	4. Rose (Bag)	0	1
Star–Ladder	5. Elephant (Glass)	0	1
Racoon–Paper	6. Truck (Arrow)	0	1
Rose–Bag	7. Insect (Acorn)	0	1
Elephant–Glass	8. Racoon (Paper)	0	1

List C	Recall C	VPAC1-8	
Rose–Bag	1. Insect (Acorn)	0	1
Racoon–Paper	2. Star (Ladder)	0	1
Star–Ladder	3. Truck (Arrow)	0	1
Reptile–Clown	4. Rose (Bag)	0	1
Elephant–Glass	5. Elephant (Glass)	0	1
Insect–Acorn	6. Reptile (Clown)	0	1
Bank–Cartoon	7. Bank (Cartoon)	0	1
Truck–Arrow	8. Racoon (Paper)	0	1

List B	Recall B	VPAB1-8	
Star–Ladder	1. Elephant (Glass)	0	1
Elephant–Glass	2. Insect (Acorn)	0	1
Insect–Acorn	3. Reptile (Clown)	0	1
Truck–Arrow	4. Rose (Bag)	0	1
Reptile–Clown	5. Star (Ladder)	0	1
Bank–Cartoon	6. Racoon (Paper)	0	1
Racoon–Paper	7. Bank (Cartoon)	0	1
Rose–Bag	8. Truck (Arrow)	0	1

List D	Recall D	VPAD1-8	
Racoon–Paper	1. Star (Ladder)	0	1
Truck–Arrow	2. Rose (Bag)	0	1
Star–Ladder	3. Insect (Acorn)	0	1
Insect–Acorn	4. Racoon (Paper)	0	1
Rose–Bag	5. Elephant (Glass)	0	1
Reptile–Clown	6. Bank (Cartoon)	0	1
Bank–Cartoon	7. Reptile (Clown)	0	1
Elephant–Glass	8. Truck (Arrow)	0	1

FAMILY PICTURES B (FAMILY NAMES – CARD 3; ALPHABET – CARD 9)

FPBT ______ ⧗

Scene A **Picnic**		**Present** **FPB1-7**			**Location** **FPB8-14**			**Activity** **FPB15-21**		
		0	1		0	1		0	1	2	
Grandmother	1.	Y	N	8.	1 2 3 4	-	15.	0	-	-	-
Grandfather	2.	N	Y	9.	1 2 4	3	16.	0	1	2	Watching dog catch frisbee
Mother	3.	Y	N	10.	1 2 3 4	-	17.	0	-	-	-
Father	4.	N	Y	11.	2 3 4	1	18.	0	1	2	BBQing hamburgers
Daughter	5.	N	Y	12.	1 3 4	2	19.	0	1	2	Throwing frisbee
Son	6.	Y	N	13.	1 2 3 4	-	20.	0	-	-	-
Dog	7.	N	Y	14.	1 2 3	4	21.	0	1	2	Catching frisbee

Scene B **Dept. Store**		**Present** **FPB22-28**			**Location** **FPB29-35**			**Activity** **FPB36-42**		
		0	1		0	1		0	1	2	
Grandmother	22.	Y	N	29.	1 2 3 4	-	36.	0	-	-	-
Grandfather	23.	N	Y	30.	1 2 3	4	37.	0	1	2	Pulling out sweater
Mother	24.	N	Y	31.	1 3 4	2	38.	0	1	2	Looking in purse
Father	25.	N	Y	32.	1 2 4	3	39.	0	1	2	Looking at shirt
Daughter	26.	N	Y	33.	2 3 4	1	40.	0	1	2	Looking in mirror
Son	27.	Y	N	34.	1 2 3 4	-	41.	0	-	-	-
Dog	28.	Y	N	35.	1 2 3 4	-	42.	0	-	-	-

BOSTON NAMING TEST (ALPHABET – CARD 9)

BNT ______ ⧗

BN1-10

		Correct without clue	**Stimulus clue**	**Phonemic clue**	**Incorrect**
1.	Bed (a piece of furniture)	3	2	1	0
2.	Pencil (used for writing)	3	2	1	0
3.	Whistle (used for blowing)	3	2	1	0
4.	Comb (used for fixing hair)	3	2	1	0
5.	Saw (used by a carpenter)	3	2	1	0
6.	Helicopter (used for air travel)	3	2	1	0
7.	Octopus (an ocean animal)	3	2	1	0
8.	Hanger (found in a closet)	3	2	1	0
9.	Camel (an animal)	3	2	1	0
10.	Pretzel (something to eat)	3	2	1	0

	BN11-30			
	Correct without clue	**Stimulus clue**	**Phonemic clue**	**Incorrect**
11. Racquet (used for sports)	3	2	1	0
12. Volcano (a kind of mountain)	3	2	1	0
13. Dart (you throw it)	3	2	1	0
14. Globe (a kind of map)	3	2	1	0
15. Beaver (an animal)	3	2	1	0
16. Rhinoceros (an animal)	3	2	1	0
17. Igloo (type of house)	3	2	1	0
18. Dominoes (a game)	3	2	1	0
19. Escalator (you go up on it)	3	2	1	0
20. Hammock (you lie on it)	3	2	1	0
21. Pelican (a bird)	3	2	1	0
22. Pyramid (found in Egypt)	3	2	1	0
23. Unicorn (mythical animal)	3	2	1	0
24. Accordion (a musical instrument)	3	2	1	0
25. Asparagus (something to eat)	3	2	1	0
26. Latch (part of a door)	3	2	1	0
27. Scroll (a document)	3	2	1	0
28. Sphinx (it's found in Egypt)	3	2	1	0
29. Trellis (used in a garden)	3	2	1	0
30. Protractor (measures angles)	3	2	1	0

MATRIX REASONING

Discontinue Rule:
4 wrong in a row, or
4 out of 5 wrong.

Reverse:
If score of 0 on Item 4 or 5, administer Items 1-3 in reverse until perfect scores on *two* consecutive items.

MART ________ ⧗

							MAR1-12	
A.	1	*2*	3	4	5	DK		
B.	1	2	3	4	*5*	DK		
C.	1	2	3	*4*	5	DK		
1.	1	2	*3*	4	5	DK	0	1
2.	1	2	*3*	4	5	DK	0	1
3.	1	2	3	*4*	5	DK	0	1
4.	1	*2*	3	4	5	DK	0	1
5.	1	2	*3*	4	5	DK	0	1
6.	*1*	2	3	4	5	DK	0	1
7.	1	2	3	4	*5*	DK	0	1
8.	*1*	2	3	4	5	DK	0	1
9.	1	*2*	3	4	5	DK	0	1
10.	1	2	3	*4*	5	DK	0	1
11.	1	2	3	4	*5*	DK	0	1
12.	*1*	2	3	4	5	DK	0	1

							MAR13-26	
13.	1	2	3	*4*	5	DK	0	1
14.	1	2	*3*	4	5	DK	0	1
15.	1	*2*	3	4	5	DK	0	1
16.	1	*2*	3	4	5	DK	0	1
17.	*1*	2	3	4	5	DK	0	1
18.	1	2	3	4	*5*	DK	0	1
19.	1	2	*3*	4	5	DK	0	1
20.	1	2	3	*4*	5	DK	0	1
21.	1	2	*3*	4	5	DK	0	1
22.	1	2	*3*	4	5	DK	0	1
23.	1	*2*	3	4	5	DK	0	1
24.	*1*	2	3	4	5	DK	0	1
25.	1	*2*	3	4	5	DK	0	1
26.	1	2	3	4	*5*	DK	0	1

VERBAL PAIRED ASSOCIATES B (OBJECT WORDS – CARD 4)

VPBT ________ ⧗

Recall	**VPB1-4**	
1. Elephant (Glass)	0	1
2. Rose (Bag)	0	1
3. Bank (Cartoon)	0	1
4. Reptile (Clown)	0	1

	VPB5-8	
5. Racoon (Paper)	0	1
6. Insect (Acorn)	0	1
7. Star (Ladder)	0	1
8. Truck (Arrow)	0	1

VERBAL PAIRED ASSOCIATES B RECOGNITION (YES/NO – CARD 1)

		VPBR1-12	
		0	1
1.	Rose–Bag	N	Y
2.	Queen–Thumb	Y	N
3.	Elephant–Glass	N	Y
4.	Baseball–Forest	Y	N
5.	Star–Ladder	N	Y
6.	Racoon–Paper	N	Y
7.	Dish–Corner	Y	N
8.	Perfume–Monkey	Y	N
9.	Truck–Arrow	N	Y
10.	Dance–Rocket	Y	N
11.	Peanut–Pencil	Y	N
12.	Bank–Cartoon	N	Y

		VPBR13-24	
		0	1
13.	Insect–Acorn	N	Y
14.	Pocket–Ribbon	Y	N
15.	Candy–Typewriter	Y	N
16.	Reptile–Clown	N	Y
17.	Wrinkle–Termite	Y	N
18.	Rose–Bag	N	Y
19.	Chicken–Submarine	Y	N
20.	Star–Ladder	N	Y
21.	Rain–Circus	Y	N
22.	Bread–Island	Y	N
23.	Elephant–Glass	N	Y
24.	Insect–Acorn	N	Y

BKBT ________ 8

BLOCK C

BABCOCK STORY RECALL TEST A

BSAT ________ ⧗

Story		**BSA1-11**	
1.	December 6.	0	1
2.	Last week	0	1
3.	a river	0	1
4.	overflowed	0	1
5.	in a small town	0	1
6.	ten miles	0	1
7.	from Albany.	0	1
8.	Water covered the streets	0	1
9.	and entered the houses.	0	1
10.	Fourteen persons	0	1
11.	were drowned	0	1

		BSA12-21	
12.	and 600 persons	0	1
13.	caught cold	0	1
14.	because of the dampness	0	1
15.	and cold weather.	0	1
16.	In saving	0	1
17.	a boy	0	1
18.	who was caught	0	1
19.	under a bridge	0	1
20.	a man	0	1
21.	cut his hands.	0	1

RE-READ STORY! ☐

DIGIT-SYMBOL CODING

SYMT ________ ⧗

Digit-Symbol Coding	☐ (2:00 + to end of 4th line)	Raw Scores: ____________	**SYMD**
Incidental Learning	☐ (No time limit)	____________	**SYMIL**
Free Recall	☐ (No time limit)	____________	**SYMFR**
Symbol Copy	☐ (1:30) ID # on all sheets ☐	____________	**SYMC**

HOOPER VISUAL ORGANIZATION TEST: COMPARISON MEASURE

VOXT ________ ⧗

	0-point response	**1-point response**	**2-point response**	**VOX1-15**		
1.			Saw	0	1	2
2.			Aeroplane	0	1	2
3.	Hatchet		Hammer	0	1	2
4.		Auto	Truck	0	1	2
5.	Fingers	Glove	Hand	0	1	2
6.	Net		Basket	0	1	2
7.	Pencil		Cane, Hockey Stick	0	1	2
8.			Teakettle	0	1	2
9.			Candle	0	1	2
10.		Animal	Cat	0	1	2
11.		Animal	Mouse, Guinea Pig, etc.	0	1	2
12.		Animal	Rabbit	0	1	2
13.		Tower, Castle	Lighthouse	0	1	2
14.			Key	0	1	2
15.	Mop		Broom	0	1	2

FINGER TAPPING TEST

FTT ________ ⧗

Dominant:___ 1. __________ 2. __________ 3. __________ 4. __________ 5.__________ **FTD1-5**
Non-Dominant 1. __________ 2. __________ 3. __________ 4. __________ 5.__________ **FTN1-5**

BABCOCK STORY RECALL TEST B

BSBT _______ ⧗

Story		**BSB1-11**
1.	December 6.	0 1
2.	Last week	0 1
3.	a river	0 1
4.	overflowed	0 1
5.	in a small town	0 1
6.	ten miles	0 1
7.	from Albany.	0 1
8.	Water covered the streets	0 1
9.	and entered the houses.	0 1
10.	Fourteen persons	0 1
11.	were drowned	0 1

		BSB12-21
12.	and 600 persons	0 1
13.	caught cold	0 1
14.	because of the dampness	0 1
15.	and cold weather.	0 1
16.	In saving	0 1
17.	a boy	0 1
18.	who was caught	0 1
19.	under a bridge	0 1
20.	a man	0 1
21.	cut his hands.	0 1

WISCONSIN CARD SORTING TEST

WCT _______ ⧗

Trials: C F N C F N

___ 1. C F N O	___ 33. C F N O	___ 1. C F N O	___ 33. C F N O
___ 2. C F N O	___ 34. C F N O	___ 2. C F N O	___ 34. C F N O
___ 3. C F N O	___ 35. C F N O	___ 3. C F N O	___ 35. C F N O
___ 4. C F N O	___ 36. C F N O	___ 4. C F N O	___ 36. C F N O
___ 5. C F N O	___ 37. C F N O	___ 5. C F N O	___ 37. C F N O
___ 6. C F N O	___ 38. C F N O	___ 6. C F N O	___ 38. C F N O
___ 7. C F N O	___ 39. C F N O	___ 7. C F N O	___ 39. C F N O
___ 8. C F N O	___ 40. C F N O	___ 8. C F N O	___ 40. C F N O
___ 9. C F N O	___ 41. C F N O	___ 9. C F N O	___ 41. C F N O
___ 10. C F N O	___ 42. C F N O	___ 10. C F N O	___ 42. C F N O
___ 11. C F N O	___ 43. C F N O	___ 11. C F N O	___ 43. C F N O
___ 12. C F N O	___ 44. C F N O	___ 12. C F N O	___ 44. C F N O
___ 13. C F N O	___ 45. C F N O	___ 13. C F N O	___ 45. C F N O
___ 14. C F N O	___ 46. C F N O	___ 14. C F N O	___ 46. C F N O
___ 15. C F N O	___ 47. C F N O	___ 15. C F N O	___ 47. C F N O
___ 16. C F N O	___ 48. C F N O	___ 16. C F N O	___ 48. C F N O
___ 17. C F N O	___ 49. C F N O	___ 17. C F N O	___ 49. C F N O
___ 18. C F N O	___ 50. C F N O	___ 18. C F N O	___ 50. C F N O
___ 19. C F N O	___ 51. C F N O	___ 19. C F N O	___ 51. C F N O
___ 20. C F N O	___ 52. C F N O	___ 20. C F N O	___ 52. C F N O
___ 21. C F N O	___ 53. C F N O	___ 21. C F N O	___ 53. C F N O
___ 22. C F N O	___ 54. C F N O	___ 22. C F N O	___ 54. C F N O
___ 23. C F N O	___ 55. C F N O	___ 23. C F N O	___ 55. C F N O
___ 24. C F N O	___ 56. C F N O	___ 24. C F N O	___ 56. C F N O
___ 25. C F N O	___ 57. C F N O	___ 25. C F N O	___ 57. C F N O
___ 26. C F N O	___ 58. C F N O	___ 26. C F N O	___ 58. C F N O
___ 27. C F N O	___ 59. C F N O	___ 27. C F N O	___ 59. C F N O
___ 28. C F N O	___ 60. C F N O	___ 28. C F N O	___ 60. C F N O
___ 29. C F N O	___ 61. C F N O	___ 29. C F N O	___ 61. C F N O
___ 30. C F N O	___ 62. C F N O	___ 30. C F N O	___ 62. C F N O
___ 31. C F N O	___ 63. C F N O	___ 31. C F N O	___ 63. C F N O
32. C F N O	64. C F N O	___ 32. C F N O	___ 64. C F N O

CATEGORIES __________ **WCA**
CORRECT __________ **WCOR**
ERRORS __________ **WERR**
PERSEVERATIVE RESPONSES __________ **WPR**
NON-PERSEVERATIVE ERRORS __________ **WNPE**
PERSEVERATIVE ERRORS __________ **WPE**

BKCT ______ ⌛

BLOCK D

FAMILY PICTURES A: COMPARISON MEASURE

FPAXT _______ ⧗

Scene C **Yard**	**Present** **FPAX1-7**			**Location** **FPAX8-14**			**Activity** **FPAX15-21**					
		0	1		0	1		0	1	2		
Grandmother	1.	N	Y	8.	2, 3, 4	1	15.	0	1	2	____________	Working in flower bed
Grandfather	2.	N	Y	9.	1, 2, 4	3	16.	0	1	2	____________	Giving the dog a bath
Mother	3.	Y	N	10.	Any	-	17.	0	-	-		-
Father	4.	Y	N	11.	Any	-	18.	0	-	-		-
Daughter	5.	Y	N	12.	Any	-	19.	0	-	-		-
Son	6.	N	Y	13.	1, 3, 4	2	20.	0	1	2	____________	Mowing the lawn
Dog	7.	N	Y	14.	1, 2, 3	4	21.	0	1	2	____________	Shaking off water

Scene D **Meal**	**Present** **FPAX22-28**			**Location** **FPAX29-35**			**Activity** **FPAX36-42**					
		0	1		0	1		0	1	2		
Grandmother	22.	N	Y	29.	2, 3, 4	1	36.	0	1	2	____________	Eating salad
Grandfather	23.	Y	N	30.	Any	-	37.	0	-	-		-
Mother	24.	N	Y	31.	1, 2, 3	4	38.	0	1	2	____________	Passing the salad bowl
Father	25.	N	Y	32.	1, 3, 4	2	39.	0	1	2	____________	Pouring salad dressing
Daughter	26.	Y	N	33.	Any	-	40.	0	-	-		-
Son	27.	N	Y	34.	1, 2, 4	3	41.	0	1	2	____________	Drinking tea
Dog	28.	Y	N	35.	Any	-	42.	0	-	-		-

DIGIT SPAN FORWARDS: COMPARISON MEASURE

DSFXT _______ ⧗

		DSFX1-10	
1.	1 – 7	0	1
2.	6 – 3	0	1
3.	5 – 8 – 2	0	1
4.	6 – 9 – 4	0	1
5.	6 – 4 – 3 – 9	0	1
6.	7 – 2 – 8 – 6	0	1
7.	4 – 2 – 7 – 3 – 1	0	1
8.	7 – 5 – 8 – 3 – 6	0	1
9.	6 – 1 – 9 – 4 – 7 – 3	0	1
10.	3 – 9 – 2 – 4 – 8 – 7	0	1

		DSFX11-16	
11.	5 – 9 – 1 – 7 – 4 – 2 – 8	0	1
12.	4 – 1 – 7 – 9 – 3 – 8 – 6	0	1
13.	5 – 8 – 1 – 9 – 2 – 6 – 4 – 7	0	1
14.	3 – 8 – 2 – 9 – 5 – 1 – 7 – 4	0	1
15.	2 – 7 – 5 – 8 – 6 – 2 – 5 – 8 – 4	0	1
16.	7 – 1 – 3 – 9 – 4 – 2 – 5 – 6 – 8	0	1

DIGIT SPAN BACKWARDS: COMPARISON MEASURE

DSBXT _______ 8

			DSBX1-14	
1.	2 – 4	(4 – 2)	0	1
2.	5 – 7	(7 – 5)	0	1
3.	6 – 2 – 9	(9 – 2 – 6)	0	1
4.	4 – 1 – 5	(5 – 1 – 4)	0	1
5.	3 – 2 – 7 – 9	(9 – 7 – 2 – 3)	0	1
6.	4 – 9 – 6 – 8	(8 – 6 – 9 – 4)	0	1
7.	1 – 5 – 2 – 8 – 6	(6 – 8 – 2 – 5 – 1)	0	1
8.	6 – 1 – 8 – 4 – 3	(3 – 4 – 8 – 1 – 6)	0	1
9.	5 – 3 – 9 – 4 – 1 – 8	(8 – 1 – 4 – 9 – 3 – 5)	0	1
10.	7 – 2 – 4 – 8 – 5 – 6	(6 – 5 – 8 – 4 – 2 – 7)	0	1
11.	8 – 1 – 2 – 9 – 3 – 6 – 5	(5 – 6 – 3 – 9 – 2 – 1 – 8)	0	1
12.	4 – 7 – 3 – 9 – 1 – 2 – 8	(8 – 2 – 1 – 9 – 3 – 7 – 4)	0	1
13.	9 – 4 – 3 – 7 – 6 – 2 – 5 – 8	(8 – 5 – 2 – 6 – 7 – 3 – 4 – 9)	0	1
14.	7 – 2 – 8 – 1 – 9 – 6 – 5 – 3	(3 – 5 – 6 – 9 – 1 – 8 – 2 – 7)	0	1

AUDITORY VERBAL LEARNING TEST A

AVAT _______ 8

List A	A1	A2	A3	A4	A5	List B	B1	A6	
Drum						Desk			Drum
Curtain						Ranger			Curtain
Bell						Bird			Bell
Coffee						Shoe			Coffee
School						Stove			School
Parent						Mountain			Parent
Moon						Glasses			Moon
Garden						Towel			Garden
Hat						Cloud			Hat
Farmer						Boat			Farmer
Nose						Lamb			Nose
Turkey						Gun			Turkey
Colour						Pencil			Colour
House						Church			House
River						Fish			River

Intrusions:

	AVI1	AVI2	AVI3	AVI4	AVI5	AVIB1	AVI6
Intrusions:							

	AVA1	AVA2	AVA3	AVA4	AVA5	AVB1	AVA6
# Correct							

FAMILY PICTURES B: COMPARISON MEASURE

FPBXT ________ 8

Scene C **Yard**	**Present** **FPBX1-7**			**Location** **FPBX8-14**			**Activity** **FPBX15-21**					
		0	1		0	1		0	1	2		
Grandmother	1.	N	Y	8.	2, 3, 4	1	15.	0	1	2	________	Working in flower bed
Grandfather	2.	N	Y	9.	1, 2, 4	3	16.	0	1	2	________	Giving the dog a bath
Mother	3.	Y	N	10.	Any	-	17.	0	-	-	________	-
Father	4.	Y	N	11.	Any	-	18.	0	-	-	________	-
Daughter	5.	Y	N	12.	Any	-	19.	0	-	-	________	-
Son	6.	N	Y	13.	1, 3, 4	2	20.	0	1	2	________	Mowing the lawn
Dog	7.	N	Y	14.	1, 2, 3	4	21.	0	1	2	________	Shaking off water

Scene D **Meal**	**Present** **FPBX22-28**			**Location** **FPBX29-35**			**Activity** **FPBX36-42**					
		0	1		0	1		0	1	2		
Grandmother	22.	N	Y	29.	2, 3, 4	1	36.	0	1	2	________	Eating salad
Grandfather	23.	Y	N	30.	Any	-	37.	0	-	-	________	-
Mother	24.	N	Y	31.	1, 2, 3	4	38.	0	1	2	________	Passing the salad bowl
Father	25.	N	Y	32.	1, 3, 4	2	39.	0	1	2	________	Pouring salad dressing
Daughter	26.	Y	N	33.	Any	-	40.	0	-	-	________	-
Son	27.	N	Y	34.	1, 2, 4	3	41.	0	1	2	________	Drinking tea
Dog	28.	Y	N	35.	Any	-	42.	0	-	-	________	-

BOSTON NAMING TEST: COMPARISON MEASURE

BNXT ________ ⌛

		BNX1-30			
		Correct without clue	**Stimulus clue**	**Phonemic clue**	**Incorrect**
1.	Tree (something that grows outdoors)	3	2	1	0
2.	House (a kind of building)	3	2	1	0
3.	Scissors (used for cutting)	3	2	1	0
4.	Flower (grows in a garden)	3	2	1	0
5.	Toothbrush (used in the mouth)	3	2	1	0
6.	Broom (used for cleaning)	3	2	1	0
7.	Mushroom (something to eat)	3	2	1	0
8.	Wheelchair (found in a hospital)	3	2	1	0
9.	Mask (part of a costume)	3	2	1	0
10.	Bench (used for sitting)	3	2	1	0
11.	Snail (an animal)	3	2	1	0
12.	Seahorse (an ocean animal)	3	2	1	0
13.	Canoe (used in the water)	3	2	1	0
14.	Wreath (a Christmas decoration)	3	2	1	0
15.	Harmonica (musical instrument)	3	2	1	0
16.	Acorn (it comes from a tree)	3	2	1	0
17.	Stilts (used to make you taller)	3	2	1	0
18.	Cactus (something that grows)	3	2	1	0
19.	Harp (a musical instrument)	3	2	1	0
20.	Knocker (it's on a door)	3	2	1	0
21.	Stethoscope (used by doctors and nurses)	3	2	1	0
22.	Muzzle (used on dogs)	3	2	1	0
23.	Funnel (used for pouring)	3	2	1	0
24.	Noose (used for hanging)	3	2	1	0
25.	Compass (for drawing)	3	2	1	0
26.	Tripod (photographers or surveyors use it)	3	2	1	0
27.	Tongs (a utensil)	3	2	1	0
28.	Yoke (used on farm animals)	3	2	1	0
29.	Palette (artists use it)	3	2	1	0
30.	Abacus (it's used for counting)	3	2	1	0

BLOCK DESIGN

Discontinue Rule:
Three consecutive scores of 0.

Reverse:
If score of 0 on either Design 5 or 6, administer Designs 1-4 in reverse until perfect score (2 points) on *two* consecutive items.

BDT ________ ⌛

BD1-9

Design	**Time Limit**	**Completion Time**	**Score**						
			0	**1**	**2**	**4**	**5**	**6**	**7**
1.	30"		0	Trial 2	Trial 1				
2.	30"		0	Trail 2	Trail 1				
3.	30"		0	Trail 2	Trial 1				
4.	30"		0	Trial 2	Trial 1				
➔ 5.	60"		0	Trail 2	Trial 1				
6.	60"		0	Trial 2	Trial 1				
7.	60"		0			16-60	11-15	6-10	1-5
8.	60"		0			16-60	11-15	6-10	1-5
9.	60"		0			21-60	16-20	11-15	1-10

BD10-14

Design	Time Limit	Completion Time	Score						
			0	1	2	4	5	6	7
10.	120"		0			36-120	26-35	21-25	1-20
11.	120"		0			66-120	46-65	31-45	1-30
12.	120"		0			76-120	56-75	41-55	1-40
13.	120"		0			76-120	56-75	41-55	1-40
14.	120"		0			66-120	46-65	36-45	1-35

AUDITORY VERBAL LEARNING TEST B

AVBT ______ ⧗

List A	A7	Recog. A		Recog. B	
Drum		Bell		Ranger	
Curtain		Hat		Pencil	
Bell		Nose		Fish	
Coffee		School		Bird	
School		Moon		Mountain	
Parent		Coffee		Towel	
Moon		River		Desk	
Garden		Curtain		Gun	
Hat		Colour		Church	
Farmer		Turkey		Boat	
Nose		Parent		Cloud	
Turkey		Farmer		Glasses	
Colour		House		Shoe	
House		Garden		Stove	
River		Drum		Lamb	

Intrusions:

	AVI7	AVI8	AVIB2
Intrusions:	☐	☐	☐

	AVA7	AVA8	AVB2
# Correct:	☐	☐	☐

BKDT ______ ⧗

APPENDIX F. INFORMATION SHEETS, CONSENT AND PAYMENT FORMS

The information sheets, consent and payment forms presented here were used in the Manawatu–Whanganui region. The clinical phase of this research was reviewed by six regional ethics committees, and due to local requirements some minor detail (such as formatting) differed between regions. Additionally, contact phone numbers for both the researchers and Health and Disability Advocacy services were localised on the information sheets of each region. Finally, the assessment time estimated for in the Information Sheet – Follow-up was varied by group. These sheets are presented at 80% of their original size.

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

INFORMATION SHEET

This research is being conducted by Duncan Babbage as part of his doctoral work at Massey University, under the supervision of Dr. Janet Leathem. They can be contacted at the Psychology Clinic, Massey University, on (06) 350 5196.

The purpose of this research is to develop techniques for the assessment of various aspects of brain functioning in people with expressive disabilities. The types of areas that are assessed include things like memory and learning, and language functioning. We are undertaking this research because at the moment there are not any good ways of assessing people who have had head injuries, and also have expressive disabilities. We think that this is an important group, and we hope through this research to improve the assessment services that can be offered to people with these sorts of disabilities.

In this research we will be assessing a group of people who have experienced a head injury, and who also have difficulties with fine motor control of their arms (things like writing and drawing) and difficulty with being able to communicate using speech. In the assessments we will do a number of activities which will involve answering questions, remembering things, and completing puzzles. These are set up so someone can do them even if they cannot write, draw or speak. We will also be assessing people who have had only a head injury, and who do not have disabilities such as described above. Finally, we will be assessing some people who have neither of these.

The responses that are made in each of these activities will be written down. This information will be coded in such a way that other people will not be able to tell who was assessed. All record forms will be stored securely or destroyed at the end of this project.

The first assessment may take up to 3 hours. This may be done in a number of sessions, and not necessarily all on the same day. If you decide at the end of these assessments that you are happy to be contacted again, at some point over the next month, we will invite you to be involved in another session. This would use similar activities to the ones done in the first session, and might again be spread over a number of sessions.

As a way of partly compensating our participants for the time they are giving to participate in this research, participants will receive $15 for this initial assessment.

You are invited to take part in this first part of the research project. If you do take part you have the right to stop at any time, and do not have to answer any questions that you do not want to. Also, you can ask any questions about the study at any time during the study. Taking part in this first part does not mean you have to participate in the follow-up assessment. Whether you take part or not will not affect any treatment you may be receiving.

If you have any queries or concerns about your rights as a participant in this study, you may wish to contact a Health and Disability Services Consumer Advocate, Advocacy Network Services Trust, on 0800 423 638.

If you wish a summary of the results of the research can be provided to you. We will send this to you at the end of the project if you request it. The research will also be presented at conferences, and published in professional journals so that other psychologists can learn from our findings. These will present group findings from our research. Information that could identify an individual would not be presented.

Thank you for your consideration,

Duncan Babbage

Janet Leathem

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

CONSENT FORM

This project has been approved by the Manawatu-Whanganui Ethics Committee. This means that the Ethics Committee may check that this study is running smoothly and that the study has followed appropriate ethical procedures. Complete confidentiality is assured. If you have any ethical concerns regarding the study, you may contact the Committee on ph/fax 0-6 356 7773.

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I understand I have the right to withdraw from the study at any time and to decline to answer any particular questions.

I agree to provide information to the researchers on the understanding that my name will not be used. *(The information will be used only for this research and publications arising from this research project).*

I agree to participate in the initial part of this study under the conditions set out in the Information Sheet.

Signed: ....

Name of person signing: ....

Name of participant: ....

Date: ....

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

CONSENT FORM — GUARDIAN

This project has been approved by the Manawatu-Whanganui Ethics Committee. This means that the Ethics Committee may check that this study is running smoothly and that the study has followed appropriate ethical procedures. Complete confidentiality is assured. If you have any ethical concerns regarding the study, you may contact the Committee on ph/fax 0-6 356 7773.

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I understand I have the right to withdraw participation from the study at any time and to decline to answer any particular questions.

I agree to provide information to the researchers on the understanding that my name will not be used. *(The information will be used only for this research and publications arising from this research project).*

I agree to allow participation in this study under the conditions set out in the Information Sheet.

Signed: ....

Name: ....

Date: ....

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

Initial Assessment

PAYMENT AND CONTACT FORM

I have received $15 in partial recognition for the time participating in this initial assessment.

YES ☐ NO ☐

I am interested in participating in the follow-up assessment in the next month. I understand that even if the researchers contact me I do not have to participate in the follow-up assessment.

YES ☐ NO ☐

I would like to receive summary information of the results of this research.

YES ☐ NO ☐

Signed: ....

Name of person signing: ....

Name of participant: ....

Researcher signature: ....

Date: ....

Contact address:

Te Kunenga ki Pūrehuroa

Inception to Infinity: Massey University's commitment to learning as a life-long journey

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

INFORMATION SHEET – FOLLOW-UP

This research is being conducted by Duncan Babbage as part of his doctoral work at Massey University, under the supervision of Dr. Janet Leathem. They can be contacted at the Psychology Clinic, Massey University, on (06) 350 5196.

This is part of research that you participated in a few weeks ago. The purpose of this research is to develop techniques for the assessment of various aspects of brain functioning in people with expressive disabilities.

In the first assessment the types of areas that were assessed included things like memory and learning, and language functioning. We are undertaking this research because at the moment there are not any good ways of assessing people who have had head injuries, and also have difficulties with fine motor control of their arms (things like writing and drawing) and difficulty with being able to communicate using speech. We think that this is an important group, and we hope through this research to improve the assessment services that can be offered to people with these sorts of disabilities.

In order to fully investigate the new assessment procedures that we are developing, it is important to see how people's performance stays the same, or changes, over time. For this reason we are contacting people who said after the last assessment they might be interested to be involved in this final part.

In these repeat assessments we will do similar things to the first assessment. This will involve a number of activities which will involve answering questions, remembering things, and completing puzzles. All of these are set up so someone can do them even if they cannot write, draw or speak.

The responses that are made in each of these activities will be written down. This information will be coded in such a way that other people will not be able to tell who was assessed. All record forms will be stored securely or destroyed at the end of this project.

This assessment will take about half the amount of time as the first assessment, which for many people will be about $1^{1}/_{2}$ – 2 hours. This may be done in a number of sessions, and not necessarily all on the same day.

As a way of partly compensating our participants for the time they are giving to participate in this research, participants will receive $15 for this follow-up assessment.

You are invited to take part in the second part of this study. The fact that you did the first assessment does not mean that you have to participate again. If you do take part you have the right to stop at any time, and do not have to answer any questions that you do not want to. Also, you can ask any questions about the study at any time during the study. Whether you take part or not will not affect any treatment you may be receiving.

If you have any queries or concerns about your rights as a participant in this study, you may wish to contact a Health and Disability Services Consumer Advocate, Advocacy Network Services Trust, on 0800 423 638.

If you wish a summary of the results of the research can be provided to you. We will send this to you at the end of the project if you request it. The research will also be presented at conferences, and published in professional journals so that other psychologists can learn from our findings. These will present group findings from our research. Information that could identify an individual would not be presented.

Thank you for your consideration,

Duncan Babbage

Janet Leathem

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6 350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

CONSENT FORM – FOLLOW-UP

This project has been approved by the Manawatu-Whanganui Ethics Committee. This means that the Ethics Committee may check that this study is running smoothly and that the study has followed appropriate ethical procedures. Complete confidentiality is assured. If you have any ethical concerns regarding the study, you may contact the Committee on ph/fax 0-6 356 7773.

I have read the Information Sheet-Follow-up and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I understand I have the right to withdraw from the study at any time and to decline to answer any particular questions.

I agree to provide information to the researchers on the understanding that my name will not be used. *(The information will be used only for this research and publications arising from this research project).*

I agree to participate in this follow-up stage of the study under the conditions set out in the Information Sheet-Follow-up.

Signed: ....

Name of person signing: ....

Name of participant: ....

Date: ....

Te Kunenga ki Pūrehuroa
Inception to Infinity: Massey University's commitment to learning as a life-long journey

Massey University
COLLEGE OF HUMANITIES & SOCIAL SCIENCES

School of Psychology
Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099
Facsimile: 64 6350 5673

Neuropsychological assessment of individuals with expressive disabilities in addition to traumatic brain injuries.

Follow-up Assessment
PAYMENT FORM

I have received $15 in partial recognition for the time participating in this follow-up assessment.

YES ☐ NO ☐

I would like to receive summary information of the results of this research.

YES ☐ NO ☐

Signed: ....

Name of person signing: ....

Name of participant: ....

Researcher signature: ....

Date: ....

Contact address:

APPENDIX G. RESPONSE CARDS

GRADED ATTENTIONAL TEST

Stimulus cards shown 70% of actual size. Each card was presented on an A6 page.

Example Card A

5

Example Card B

5 – 3

Card 1

4

Card 2

4 or 9

Card 3

4 or 9 or 6

Card 4

4 – 9

Card 5

$4 - 9$, or $6 - 2$

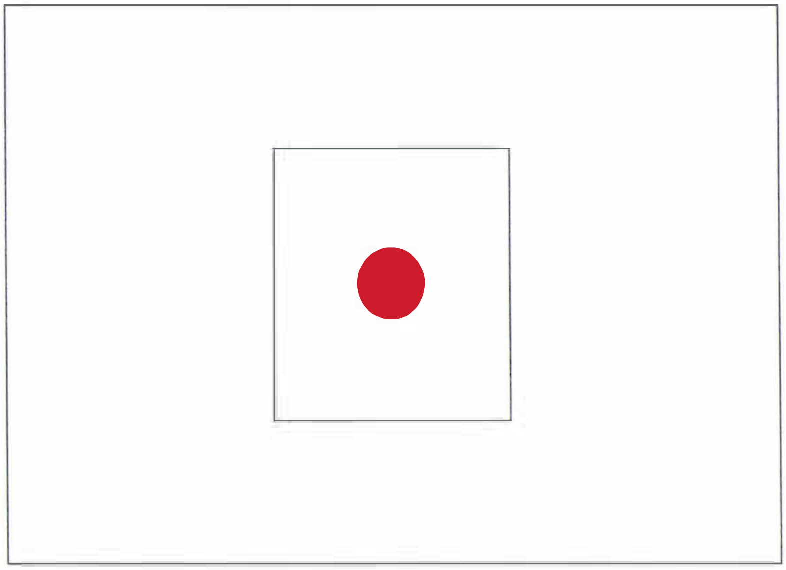
Card 6

$4 - * - 9$

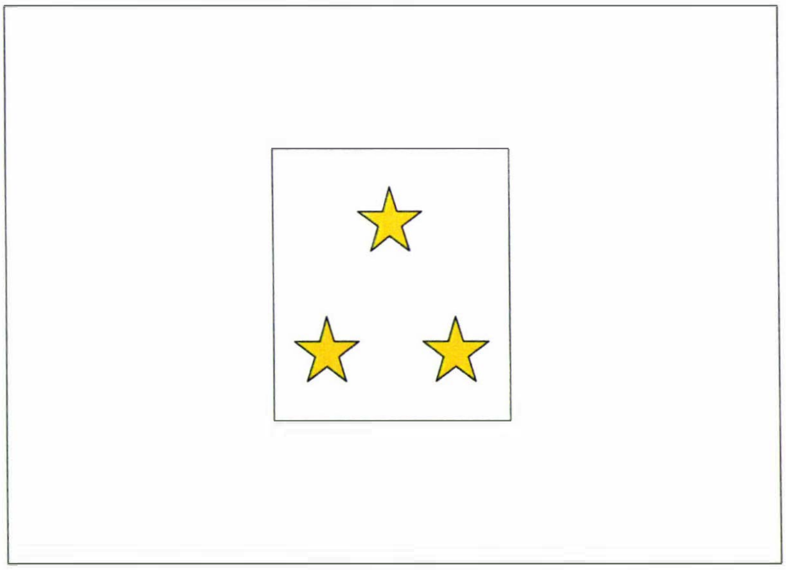
MATCH AND SHIFT CATEGORIES TEST

Stimulus cards shown 45% of actual size. Each card was presented on an A4 page.

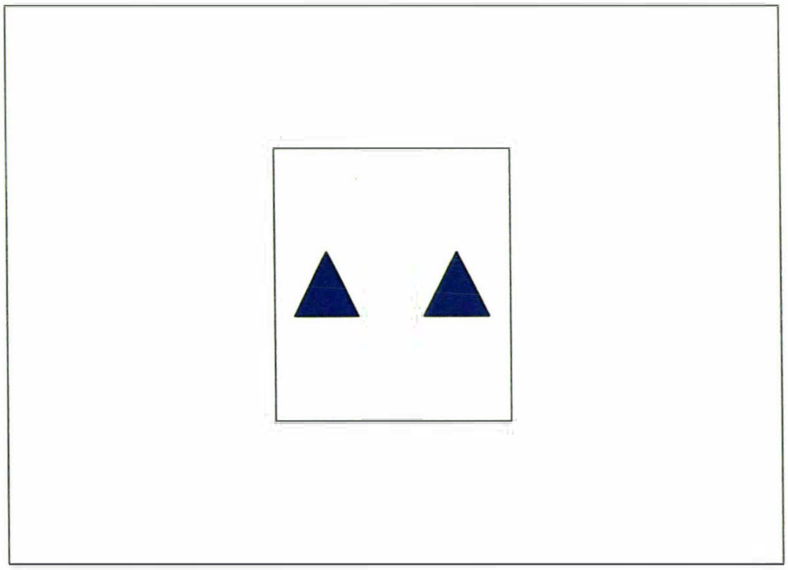
Stimulus Card 1



Stimulus Card 2



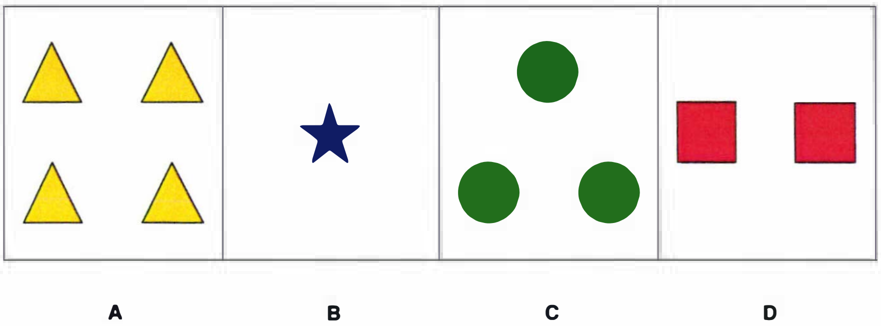
Stimulus Card 3



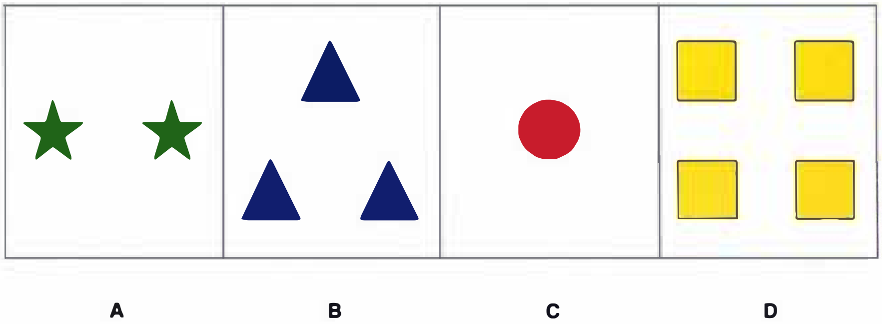
MATCH AND SHIFT CATEGORIES TEST

Response cards shown 45% of actual size. Each card was presented on an A4 page.

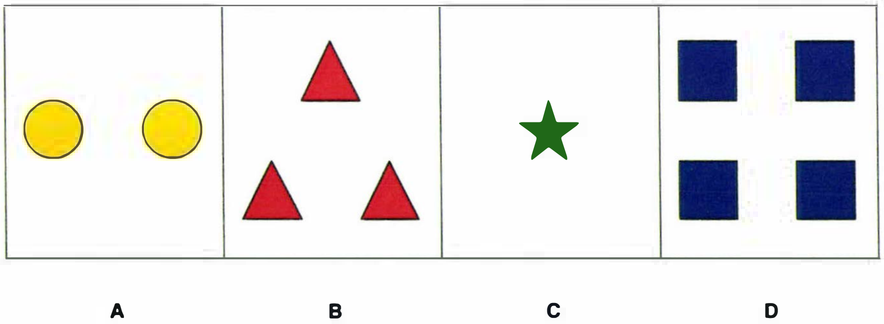
1



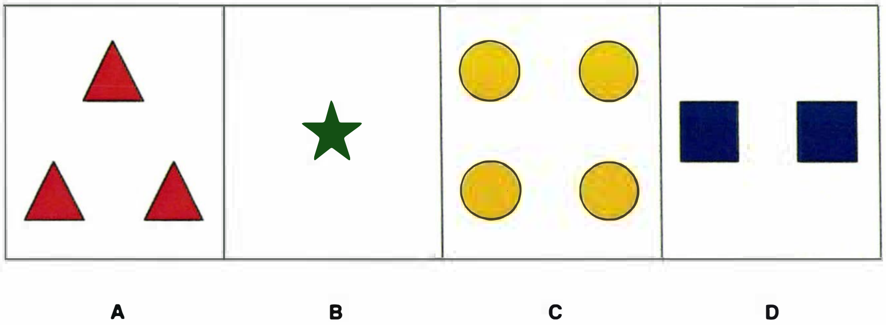
2



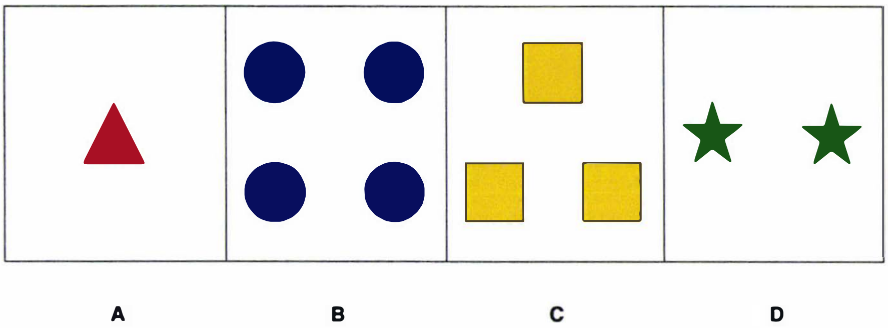
3



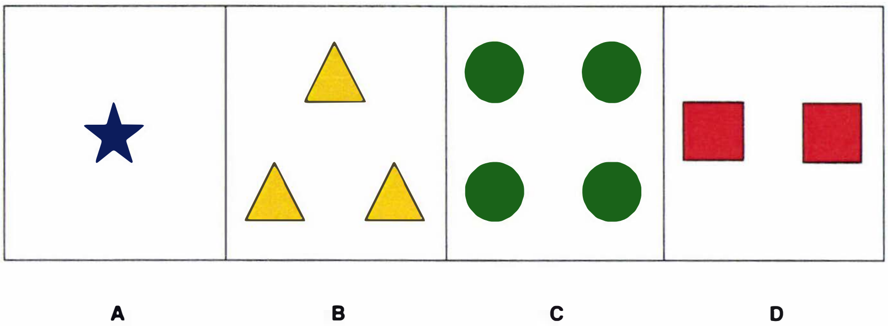
4



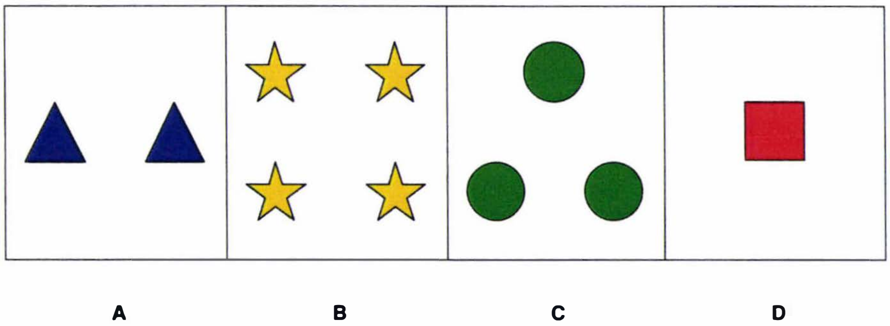
5



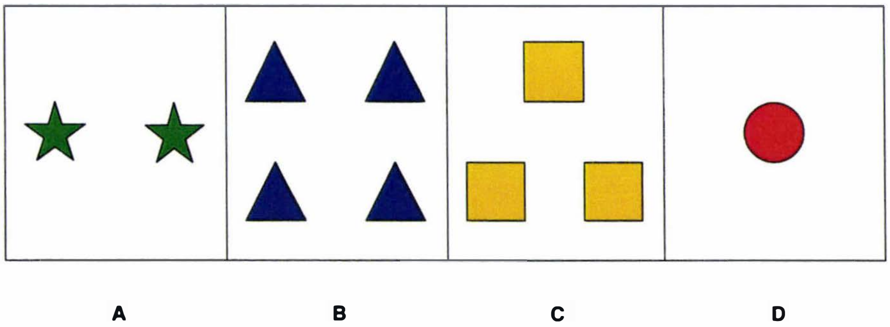
6



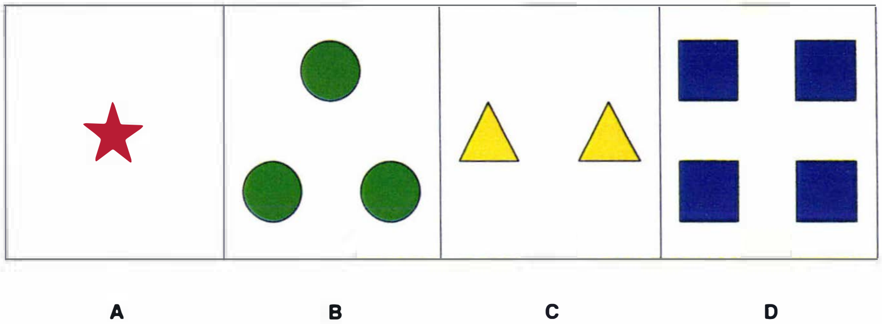
7



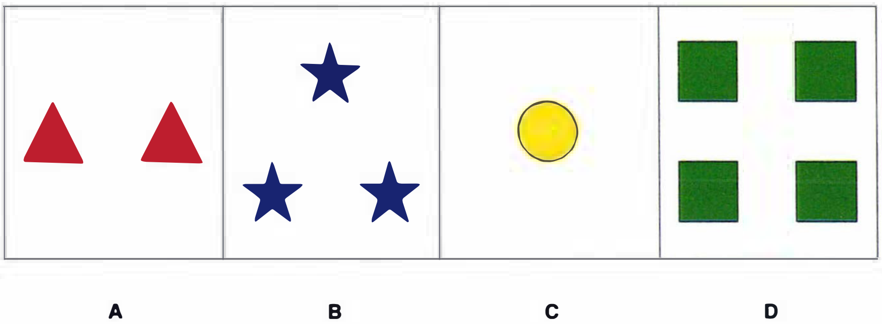
8



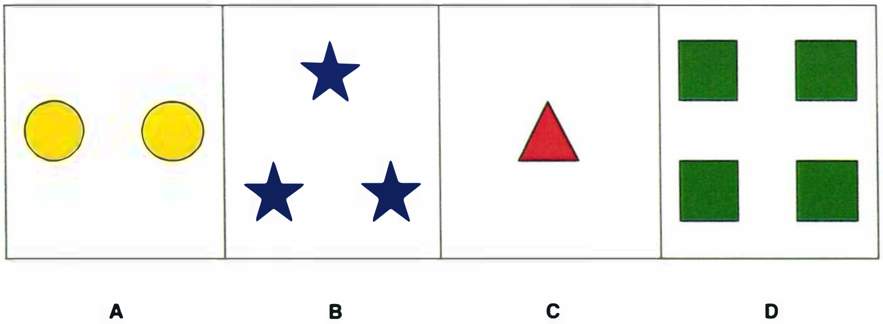
9



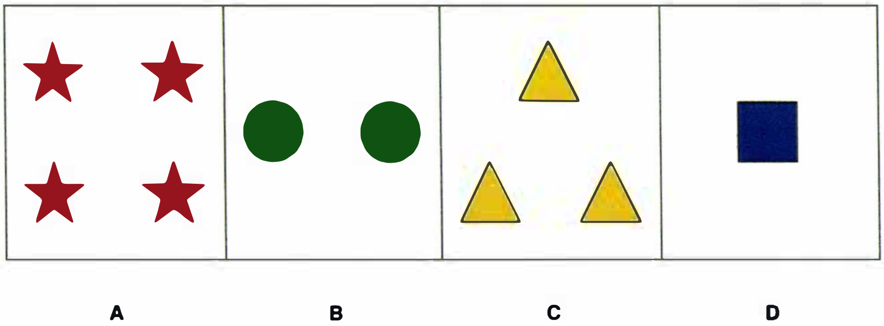
10



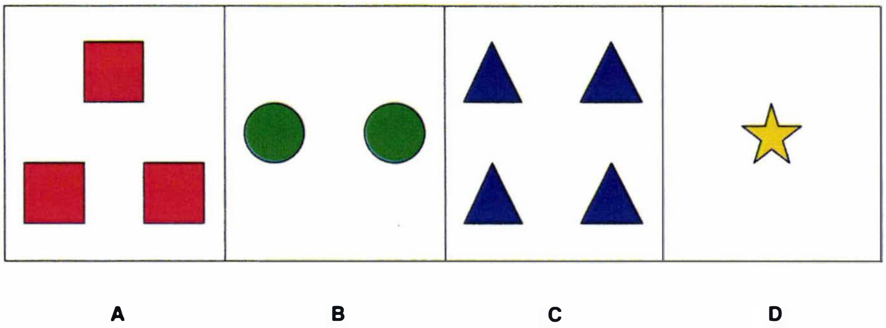
11



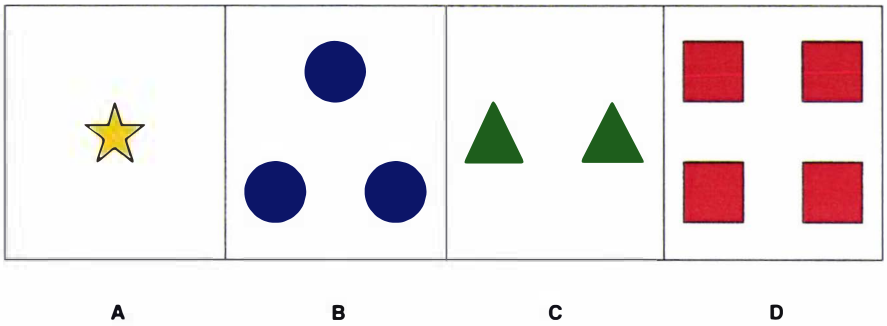
12



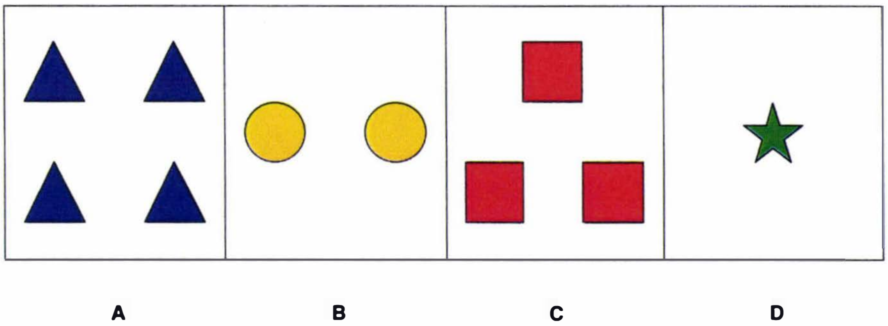
13



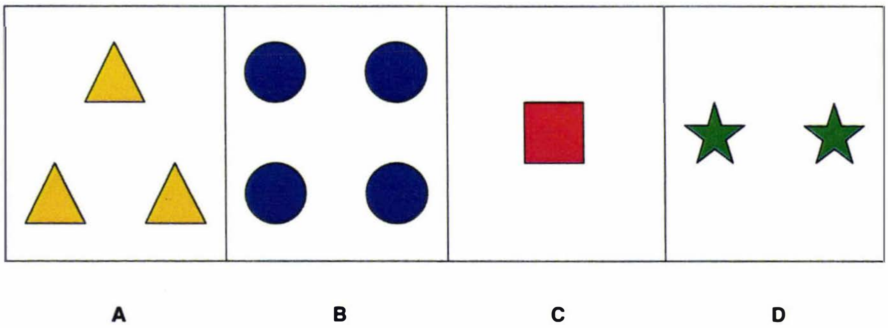
14



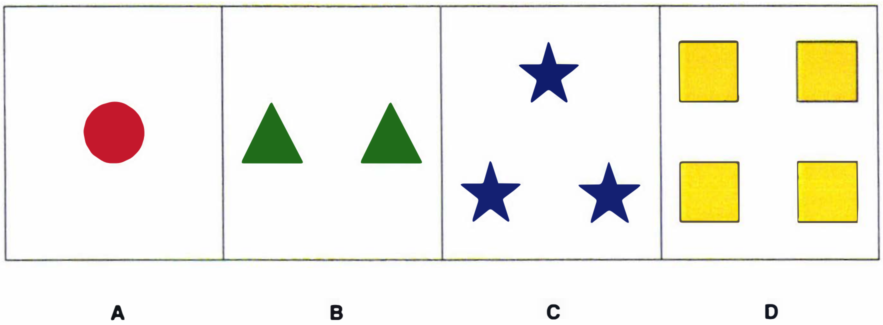
15



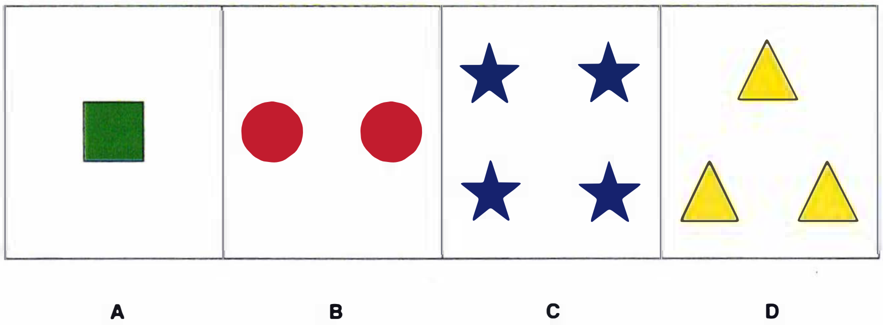
16



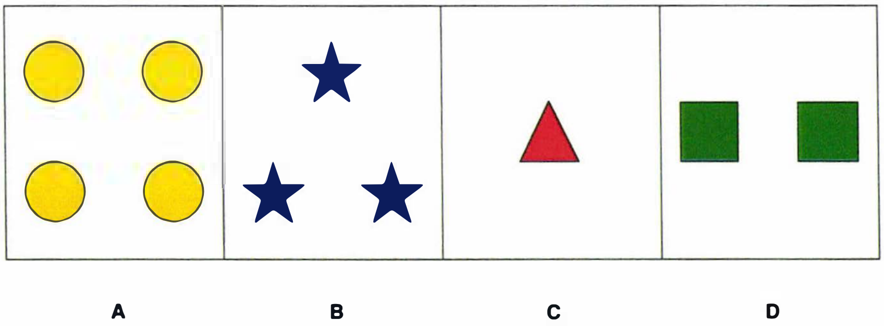
17



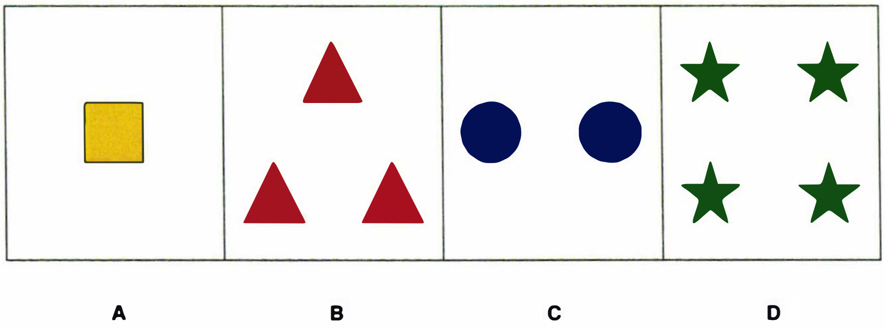
18



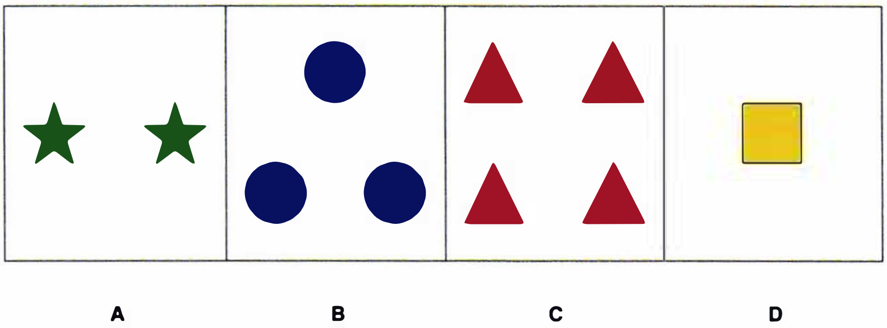
19



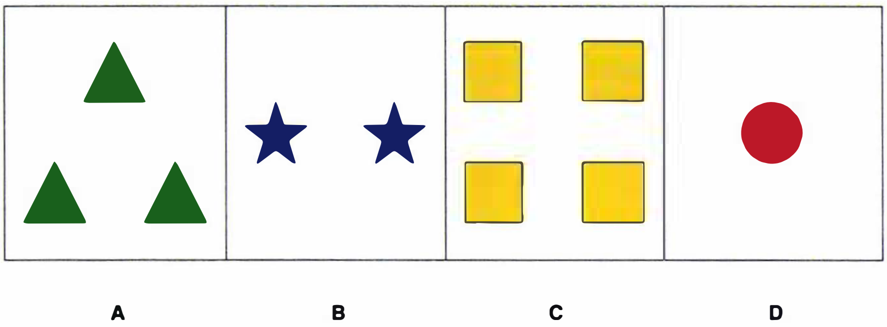
20



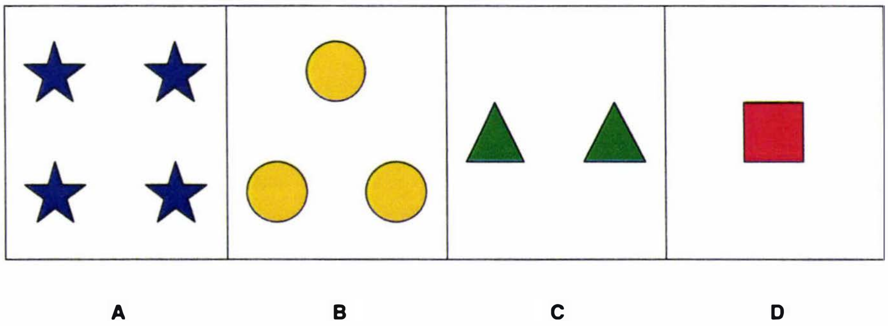
21



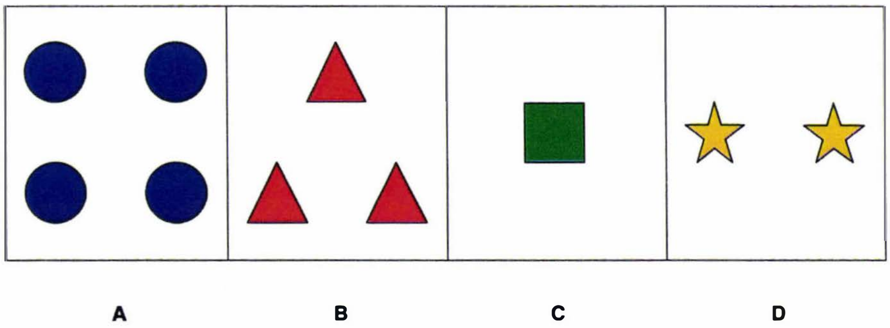
22



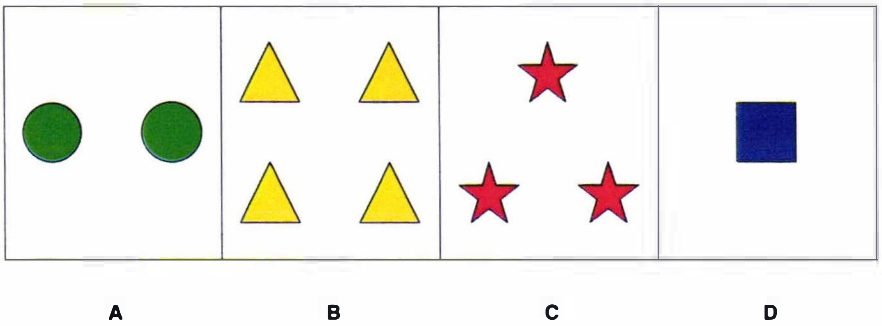
23



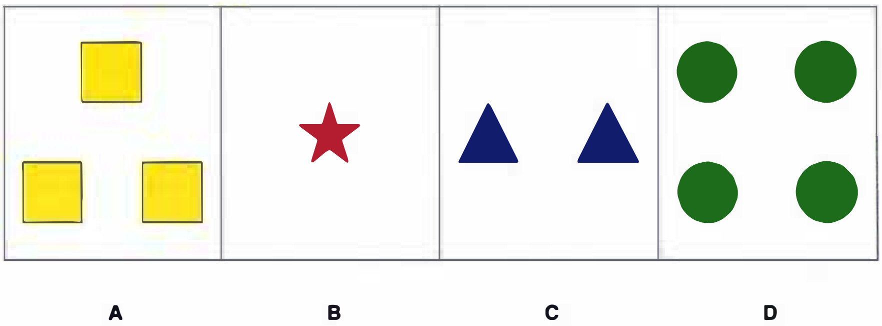
24



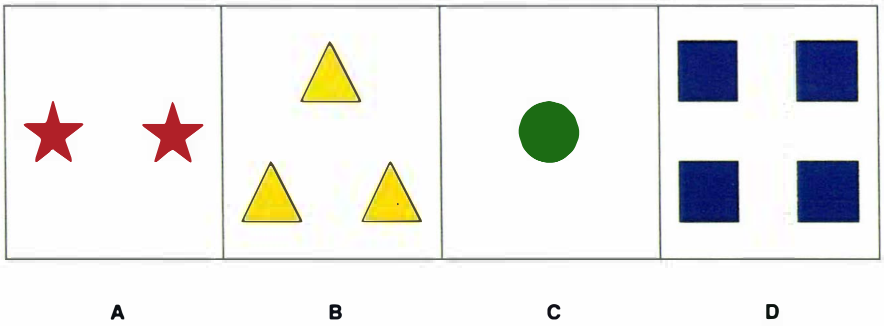
25



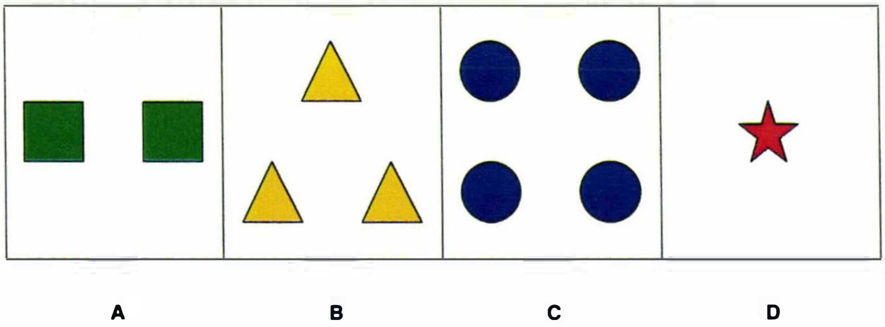
26



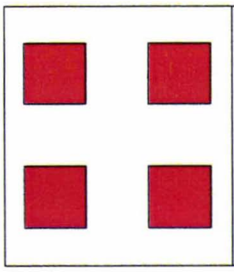
27

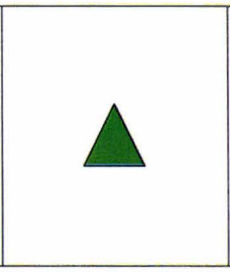


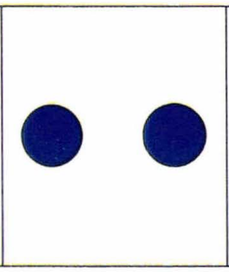
28

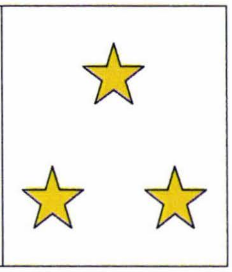


29

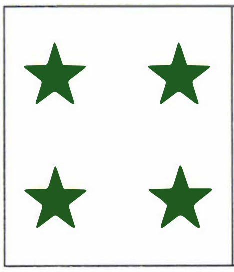

A

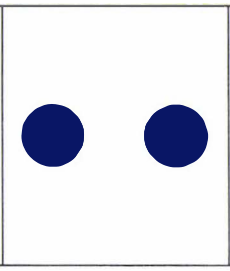

B

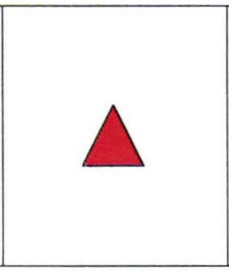

C

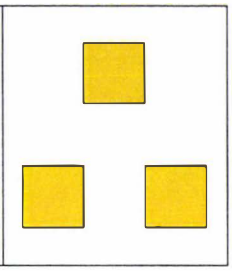

D

30


A


B


C


D

YES

NO

I DON'T
KNOW

I MADE
A MISTAKE

1	**2**	**3**
4	**5**	**6**
7	**8**	**9**
	0	

Yes	**No**	**I Don’t Know**	**I Made A Mistake**

GRANDFATHER GRANDMOTHER

FATHER MOTHER

SON DAUGHTER

DOG

YES NO I DON'T KNOW I MADE A MISTAKE

Arrow

Clown

Ladder

Bag

Acorn

Cartoon

Paper

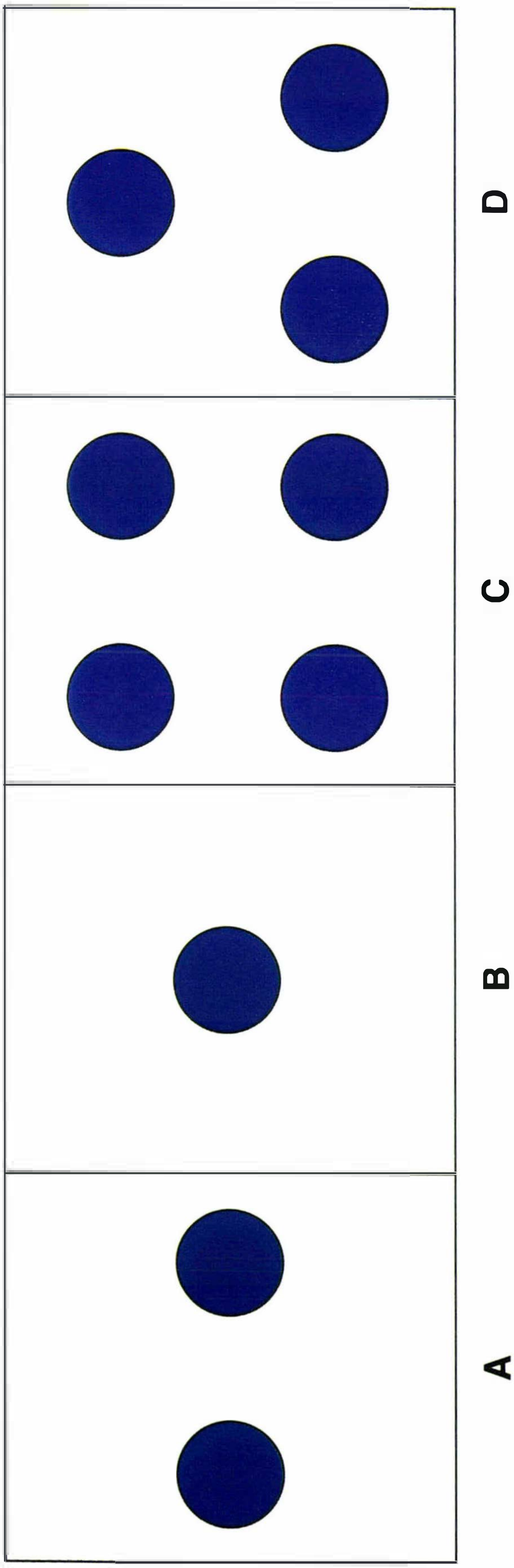
Glass

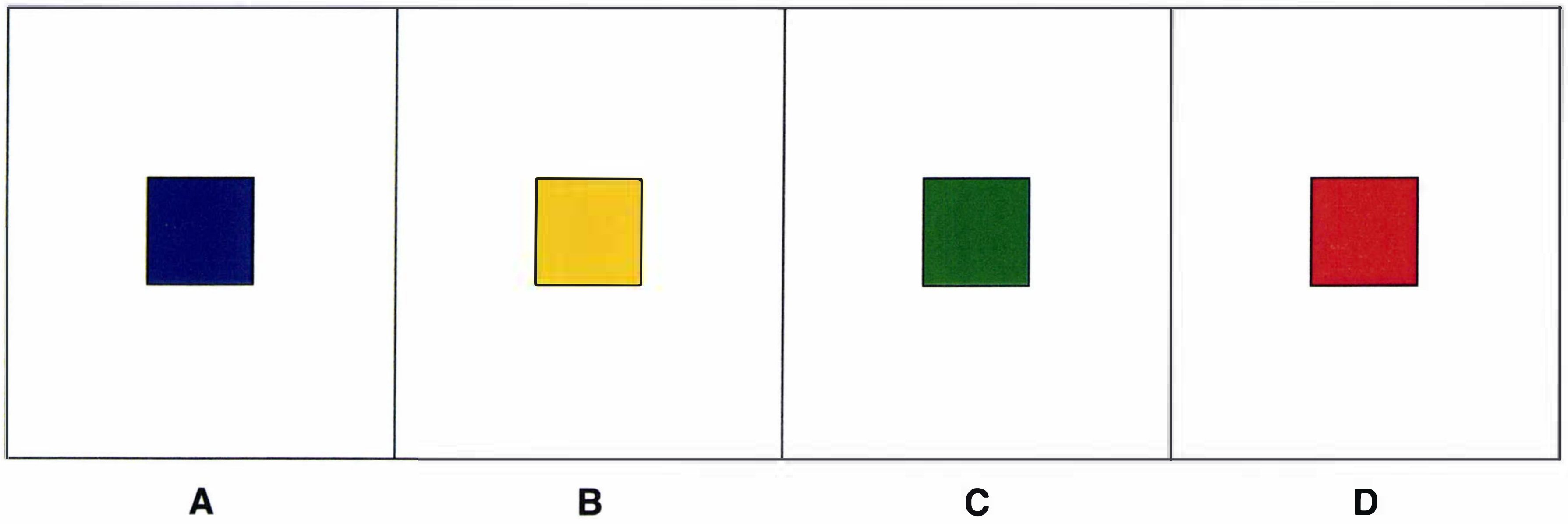
Yes

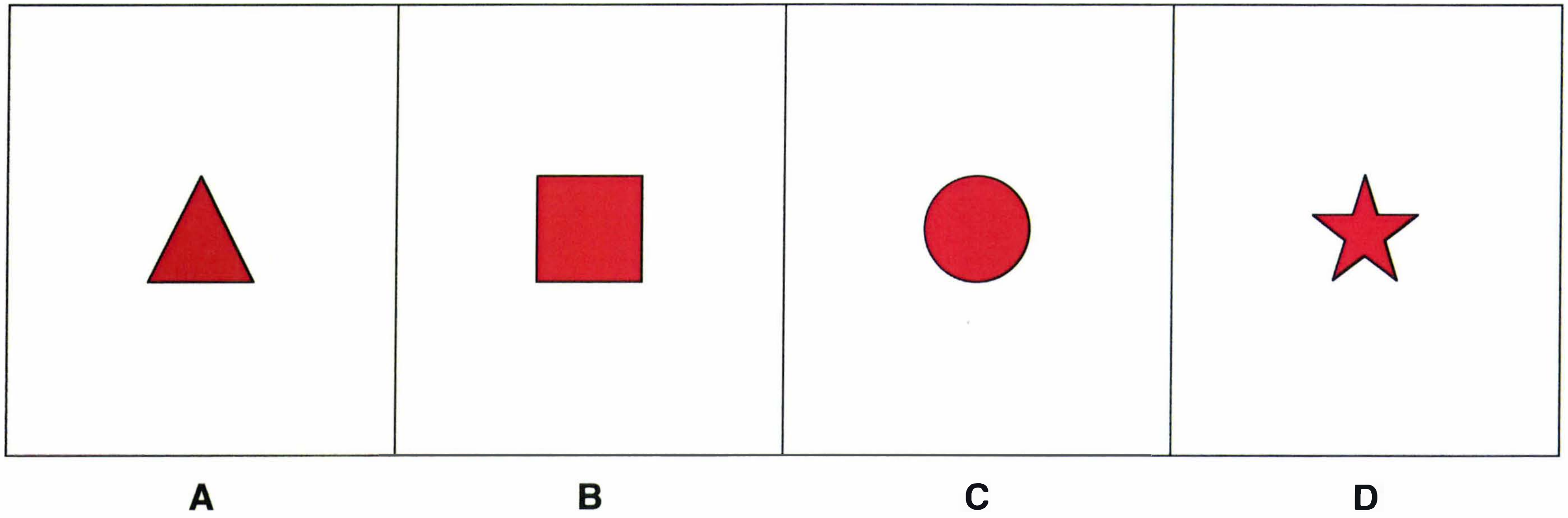
No

I Don't Know

I Made A Mistake


D
C
B
A


A
B
C
D


A
B
C
D

PAKEHA/
NZ EUROPEAN

MAORI

PACIFIC ISLANDER

ASIAN

OTHER EUROPEAN

OTHER

YES

NO

I DON'T
KNOW

I MADE
A MISTAKE

A B C D E F

G H I J K L

M N O P Q R

S T U V W X

Y Z YES NO THE AND

AVLT WORD LIST RECOGNITION

BELL
WINDOW
HAT
BARN
RANGER
NOSE
WEATHER
SCHOOL
HAND
PENCIL

HOME
FISH
MOON
TREE
BALLOON
BIRD
MOUNTAIN
COFFEE
MOUSE
RIVER

TOWEL
CURTAIN
FLOWER
COLOUR
DESK
GUN
CRAYON
CHURCH
TURKEY
FOUNTAIN

BOAT
HOT
PARENT
WATER
FARMER
ROSE
CLOUD
HOUSE
STRANGER
GARDEN

GLASSES
STOCKING
SHOE
TEACHER
STOVE
NEST
CHILDREN
DRUM
TOFFEE
LAMB

APPENDIX H. BABCOCK STORY RECALL TEST SCORING PROCEDURES

The criteria below were used in the clinical phase for scoring both the initial and delayed recall trials of the Babcock Story Recall Test. One point was awarded for each section of the story which was correct, repeating either the original words, or any of the acceptable alternatives listed below. Following the instructions in Lezak (1995), four points were added to initial recall scores to compensate for the re-reading of the story before the recall delay.

The principle underlying the selection of the acceptable alternative responses was that reasonable synonyms for the words and meaning of the original story was acceptable, and that correct material did not have to be presented in the correct context in order to get the marks allocated. Reasonable synonyms other than the examples provided here were also accepted.

Original Story	**Alternative Acceptable Responses**
December 6.	Any variation of *December 6.*
Last week	*A week ago; Seven days ago.*
a river	*River* in any context.
overflowed	Any indication of *flooding* in any context.
in a small town	Any indication of a small town.
ten miles	*Ten miles* only.
from Albany.	An indication of events at a distance from Albany (not *in Albany* or simply *Albany*).
Water covered the streets	*Water flowed through the streets.*
and entered the houses.	*Water filled the houses*, or similar.
Fourteen persons	*Fourteen persons* in any context.
were drowned	Any indication of death by drowning.
and 600 persons	An indication of 600 people in any context.
caught cold	*Caught cold* only.
because of the dampness	Indication of sickness due to dampness.
and cold weather.	Indication of sickness due to cold weather.
In saving	*Rescuing.*
a boy	Qualifiers acceptable, e.g., *young boy; 6-year-old boy.*
who was caught	An indication a person was caught or stuck.
under a bridge	*Under a bridge* in any context.
a man	An indication of an adult male in any context.
cut his hands.	*Cut his hand; cut his arm(s).*

APPENDIX I. TEST ADMINISTRATION AND DELAYED RECALL TIMES

Table A1. Descriptive statistics for test administration times in minutes, by group.

	Mean	Std Dev.	Minimum	Median	Maximum
Good Samaritan Hospital Orientation Test					
ED	1.90	0.82	1.28	1.63	3.87
TBI	1.54	0.75	0.73	1.39	3.53
Norm	1.76	1.47	0.48	1.25	8.00
Graded Attentional Test					
ED	6.85	1.43	4.63	6.63	9.00
TBI	6.34	1.24	5.17	5.78	8.98
Norm	5.81	1.07	3.82	5.49	9.47
Digits Forwards and Backwards					
Digits Forwards					
ED	3.74	1.95	1.17	3.35	7.08
TBI	3.54	0.95	1.90	3.57	5.63
Norm	3.39	0.98	1.72	3.32	7.00
Digits Backwards					
ED	3.12	1.15	1.23	2.96	5.08
TBI	3.39	1.36	1.33	3.08	6.40
Norm	3.29	1.39	1.13	3.12	9.00
Digits Forwards comp.					
TBI	2.39	0.77	1.23	2.18	4.67
Norm	2.33	0.79	0.78	2.20	4.30
Digits Backwards comp.					
TBI	2.01	0.82	0.77	1.95	3.73
Norm	2.01	0.86	0.48	1.91	4.58
Logical Memory and Babcock Story Recall					
Logical Memory A					
ED	5.52	1.90	2.78	4.85	8.53
TBI	4.27	1.18	2.70	4.27	6.93
Norm	3.83	1.20	1.83	3.47	9.15
Logical Memory B					
ED	3.74	3.77	1.50	2.27	13.00
TBI	2.07	0.59	1.40	2.01	3.17
Norm	1.86	0.73	0.88	1.68	5.00
Babcock Story Recall A					
TBI	2.92	0.86	1.48	2.68	4.73
Norm	2.70	0.87	0.98	2.52	5.38

(table continues)

	Mean	Std Dev.	Minimum	Median	Maximum
Babcock Story Recall B					
TBI	1.24	0.47	0.65	1.05	2.50
Norm	1.43	0.70	0.38	1.26	3.43
Verbal Paired Associates and Auditory Verbal Learning Test					
Verbal Paired Associates A					
ED	15.20	5.76	8.72	13.45	25.35
TBI	8.53	1.58	5.67	8.48	11.83
Norm	7.49	1.81	4.00	7.28	14.40
Verbal Paired Associates B					
ED	5.13	2.84	2.58	4.02	12.13
TBI	2.67	0.65	1.53	2.58	4.08
Norm	2.21	0.58	1.33	2.07	4.00
Auditory Verbal Learning Test A					
TBI	8.06	2.05	5.87	7.47	14.27
Norm	7.40	1.98	3.97	6.92	14.65
Auditory Verbal Learning Test B					
TBI	3.17	0.65	1.87	3.08	4.43
Norm	3.23	0.86	1.47	3.09	6.07
Faces					
Faces A					
ED	9.39	6.11	4.70	7.28	25.27
TBI	5.05	1.33	3.42	4.59	7.73
Norm	5.38	1.49	2.73	5.05	12.50
Faces B					
ED	6.16	5.17	2.73	4.77	18.53
TBI	3.08	0.75	2.13	2.92	4.82
Norm	3.81	1.18	2.42	3.45	8.00
Family Pictures					
Family Pictures A					
ED	11.31	4.83	6.33	10.58	23.45
TBI	8.14	2.43	4.37	7.87	14.68
Norm	8.17	2.51	4.00	7.88	20.00
Family Pictures B					
ED	10.44	6.62	3.07	8.78	25.37
TBI	4.86	1.94	2.32	4.33	8.65
Norm	4.84	1.62	2.85	4.53	10.88
Family Pictures A comp.					
TBI	5.10	1.36	3.27	4.98	8.52
Norm	4.90	1.42	2.48	4.75	10.00
Family Pictures B comp.					
TBI	2.98	1.33	1.72	2.57	7.37
Norm	3.11	1.10	1.68	2.88	9.52
Auditory Reception					
ED	6.37	4.61	2.77	4.28	15.40
TBI	3.06	0.61	2.23	2.97	4.45
Norm	3.21	1.20	1.67	2.73	8.07

(table continues)

	Mean	Std Dev.	Minimum	Median	Maximum
Boston Naming Test					
Boston Naming Test					
ED	17.87	9.23	8.30	18.25	33.07
TBI	9.46	2.36	5.08	9.20	14.08
Norm	7.99	2.61	4.32	7.24	15.00
Boston Naming Test comp.					
TBI	4.02	1.27	2.12	3.70	6.93
Norm	2.95	1.25	1.15	2.67	6.45
Familiar and Novel Language Comprehension test					
ED	17.39	9.77	10.20	14.70	42.27
TBI	8.24	1.99	5.80	7.93	12.63
Norm	7.92	2.09	4.93	7.08	12.55
Hooper Visual Organization Test					
Hooper Visual Organization Test					
ED	6.91	2.64	3.60	6.88	10.70
TBI	5.40	1.87	3.38	5.08	11.08
Norm	5.45	2.12	2.68	4.94	14.43
Hooper Visual Organization Test comp.					
TBI	2.21	0.72	1.33	2.23	3.85
Norm	2.30	1.06	0.62	2.06	5.37
Block Design					
TBI	12.33	2.59	7.05	12.98	18.05
Norm	10.66	2.88	5.65	10.40	19.05
Matrix Reasoning					
ED	7.92	6.17	2.87	5.87	23.23
TBI	8.30	4.43	1.60	7.98	19.47
Norm	9.94	4.46	2.80	9.20	25.12
Match and Shift Categories test and Wisconsin Card Sorting Test					
Match and Shift Categories test					
ED	14.70	5.08	8.07	13.87	23.18
TBI	7.15	2.56	3.47	6.52	14.18
Norm	6.25	1.95	3.28	6.00	13.00
Wisconsin Card Sorting Test					
TBI	12.69	4.09	6.47	12.07	19.95
Norm	9.63	3.34	4.72	8.35	19.00
Digit Symbol					
TBI	9.15	1.83	5.28	8.93	12.47
Norm	8.18	1.55	3.75	7.87	13.17
Finger Tapping					
TBI	4.76	0.87	3.50	4.85	6.85
Norm	4.81	1.14	1.75	4.76	8.33

Table A2. Descriptive statistics for delayed recall times in minutes, by group.

	Mean	**Std Dev.**	**Minimum**	**Median**	**Maximum**
Logical Memory					
ED	25.21	8.86	15.23	25.27	42.27
TBI	18.21	4.97	9.50	17.95	30.67
Norm	19.17	5.64	5.88	17.55	46.15
Differences:	F(2, 104)=4.841, p=.010 $\chi^2(2)=5.361$, p=.069				
Babcock Story Recall					
TBI	16.17	3.61	10.70	16.63	22.65
Norm	15.12	2.50	7.87	15.03	24.63
Differences:	F(1, 94)=2.075, n.s.				
Verbal Paired Associates					
ED	22.60	12.19	5.40	22.31	44.02
TBI	22.62	6.89	11.77	21.93	38.00
Norm	22.77	6.61	5.73	22.06	42.83
Differences:	F(2, 103)=0.005, n.s.				
Auditory Verbal Learning Test					
TBI	20.29	5.61	13.02	18.55	38.87
Norm	16.72	3.96	10.13	16.30	28.98
Differences:	F(1, 90)=9.579, p=.003				
Faces					
ED	19.05	7.78	9.12	18.85	30.93
TBI	11.47	2.15	8.93	10.43	15.98
Norm	10.91	2.56	3.55	10.36	19.00
Differences:	F(2, 104)=25.801, p<.001				
Family Pictures					
ED	22.93	12.28	5.60	20.75	44.45
TBI	16.85	6.88	10.87	14.68	41.75
Norm	14.14	2.70	9.48	14.07	25.00
Differences:	F(2, 103)=13.430, p<.001				
Family Pictures comp.					
TBI	12.54	2.80	8.78	11.92	20.42
Norm	12.03	3.35	5.60	11.02	21.97
Differences:	F(1, 93)=0.346, n.s.				

APPENDIX J. NORMATIVE DATA COMPUTATION PROCEDURE

Purpose of the Procedure

It was decided to convert scores observed in the normative sample into standardised scores, suitable for use by clinicians as preliminary normative data for the adapted measures described in the clinical phase of this research. The following procedure was used to develop standard score conversion tables for each of the measures. These conversion tables allow any future users of the tests to easily convert the raw scores into standardised scores, by consulting tables of a familiar style. The standardised scores were based on a theoretical normal distribution, with a range from 1 to 19, a mean of 10, and a standard deviation of 3. The actual conversion tables are presented in Table 26 (*Results*, p. 188).

While the aim was to develop standard scores in the format of those used in the Wechsler scales, the procedure described below was developed independently. Nonetheless, on the basis of the brief information presented in the Wechsler technical manual (Wechsler, 1997b), the procedure appears to be essentially equivalent. The key difference between the two approaches is that the current procedure utilised a solely mathematical approach, while for the Wechsler scales an additional step was undertaken of manually 'smoothing' the distributions to bring means, standard deviations, and skewness closer to the desired distributions.

This appendix presents in detail the procedure used to develop the standardised scores from the normative group. The method that was used to identify the standard score which corresponds to each observed raw score is described. Following this, the procedure is described which was used to handle cases where potential scores on the scale were not observed in the normative group. This is a particular issue with the current small normative sample, and is largely resolved if the sample size is increased, as all possible scores on a measure are likely to be observed. Finally, other issues which limit the conclusions that can be drawn from these norms are discussed.

Procedure for Standard Score Computation

First, the range of each standard score was calculated in z-score units. For example, a standard score of 7 should be one standard deviation below the mean. However, this score in fact represents the midpoint of a continuous range of the normal distribution. Each standard score includes ± $^1/_6$ of a standard deviation around its midpoint. That is, a standard score of 7 encompasses points on the normal distribution with z-scores of -1.17 through -0.84; a standard score of 10 includes all points with z-score values between ± 0.17.

For each of these z-score ranges, associated percentile ranges were calculated from a table of areas under the theoretical normal distribution (Coolican, 1994). A standard score of 7 thus represents the range from the 12.10th percentile to the 20.05th percentile of the normal distribution, while a score of 10 covers the 43.25th percentile to the 56.75th percentile.

Using these percentile ranges from the normal distribution, raw scores which fell within each of these ranges in the observed normative group distribution were identified[39]. In order to do this, observed values were entered into a distribution analysis as though they were midpoints of observed groups[40]. This procedure interpolates missing values down to the next lower observed value, and up to the next higher observed value across the percentile range that the original raw data point occupied, while holding the original midpoint position constant in the percentile distribution. In fact, the end result of this procedure is mathematically equivalent to calculating the midpoint simply by computing the mean value of the range: (Bottom of range + Top of range) / 2. The advantage of the more complicated procedure lies in the fact that it also interpolates where a particular data value should fall, in cases where such a value did not occur in the observed distribution.

---

[39] This is the only step in the process where additional error is introduced; unfortunately, the amount of error that can be introduced is substantial. See Chapter 9, *Discussion,* for an elaboration of these issues.

[40] This was done using the '/GROUPED' subcommand while computing percentiles with the 'FREQUENCIES' command in SPSS. While this command may not have been designed with this use in mind, the mathematical effect is as described, and is appropriate for the current purpose.

This apparently complex, but actually mathematically simple, process is illustrated in Figure A1, which displays the bottom 10 percent of the distribution of a hypothetical variable (based on the observed distribution of Familiar and Novel Language Comprehension Test Familiar Language scores from the first 80 participants in the normative group).

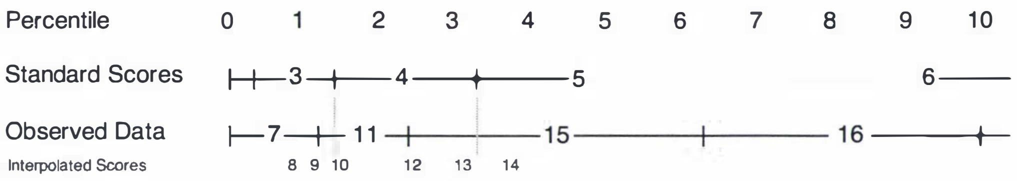


Figure A1. Conversion of observed and interpolated data to standard scores.

The figure shows the percentile range covered by the standard scores 3 through 5, and part of the range covered by 6 (the upper bound being the 12.10th percentile). Displayed also in the figure are the observed values for the variable. The individual who achieved a raw score 7 on the test scored the lowest in the sample (bottom 1.25% of the sample); the next individual scored 11 (next 1.25%), and the following 3 individuals all achieved raw scores of 15 (3.75% of the sample).

For each raw score value, the midpoint of the distribution was calculated, and the standard score range in which it fell was identified. In the figure above, the midpoint of the percentile distribution of raw score of 7 falls in the range of standard score 3, and so 7 is allocated this standard score. Similarly, raw score 11 becomes standard score 4, raw score 15 becomes 5, and 16 becomes standard score 6.

The figure also displays the interpolated scores which were not observed in the initial sample (raw scores of 8–10, 12–14). The procedure described above fixes for each observed value the mid-point on the distribution. Following this, the missing data points are interpolated linearly between the adjacent variables. As illustrated in the figure, this results in values 8–10 being spaced much more closely that the values 12–14 through the percentile range.

Once these values have been interpolated, they are allocated to standard scores in the same manner as the observed values. In the example displayed in the figure, 8 and 9 become standard score 3, while a value of 10 just falls into the range of standard score 4, along with 12 and 13. Raw score 14 becomes a standard score of 5.

This procedure was completed for all variables, and the results tabulated in Table 26. Values which fell outside of the observed range in the normal distribution were allocated to the most extreme standard score for the observed distribution. In the example above, all values ≤ 7 would be allocated standard score 3. This procedure was considered more appropriate than extrapolation, since it no information was available on the shape of the distribution beyond the most extreme observed values.

Discussion of the Basis of the Computation Procedure

The current procedure was a reasonably effective way to produce standard score conversion tables from a relatively small normative data set, and where there were valid variable values which were not observed in the sample. Examination of the fit of these standardised scores to the theoretical normal distribution on which they were based is presented in Chapter 8, *Results*.

There were clearly inadequacies to this procedure. The interpolation of values not observed may be theoretically sensible but raises issues of validity of the normative information. However, the use of normative information based on such a small sample should *always* be evaluated carefully, and such results used as general indications rather than fine-grained analysis. Nonetheless, the current procedure represents the most mathematically justifiable response to the need to allocate the values which were not observed to a standardised score.

A further difficulty with the current procedure occurs where a sample is not simply skewed, but comes from a markedly non-normal distribution. This was observed in a number of variables in the current research, where due to a low ceiling the majority of participants in the normative group score at the top, or near the top, of the scale. As all of these individuals have the same raw scores, all must be allocated to the same standardised score. For the variable described in the figure above, only one of a number of such extreme examples, almost 49% of the sample achieved a perfect score. This range of the distribution thus overlaps with parts of the distribution allocated to standard scores from 10 through to 19. The same procedure was used here as for other values, allocating a standard score of 12 on the basis of the group midpoint[41]. Nonetheless, the limitations of this procedure are clear.

Where there were valid raw data points not observed in the sample, a more sophisticated interpolation approach would have been to use curvilinear interpolation. Such a procedure would attempt to fit a curve which passed through all observed data points in the distribution (i.e., 7, 11, and 15 above), placing the interpolated points at intervals which matched the rate of change in distance between the observed variables. However such a procedure, while theoretically possible, was not pursued for a number of reasons. First, there is no clear evidence to suggest that this would produce more valid standard scores than a linear interpolation. Secondly, it is considerably more complex computationally, for little likely practical return. Thirdly, and most importantly, it assumes highly complex relationships, which in the current research would have been based on data points at times representing only single observations (such as the values of 7 and 11 in the figure above).

The size of the sample also places boundaries on the range of standardised scores which will can be matched to observed data. Regardless of the range of actual observed values, data are fitted to a normal curve. Standardised scores of 1 and 2 together account for only 0.62% of the normal distribution. Therefore, for the score of just one individual to fall within this combined range would require a sample size greater than $^1/_{0.0062}$, or at least 162 individuals. For the score of one individual to fall in the range of standard score 1 would require a sample of $^1/_{0.0023}$, or 435. The same figures apply to standard scores 18 and 19. Therefore, with the sample size of the current project, the midpoints of all observed scores could only fall within the ranges allocated to standard scores 3 through 17, bounds which are indeed observed in the conversion table (Table 26, *Results*).

---

[41] Recall that standard scores 13 through 19 account for a smaller percentage of the normal distribution than scores 10 though 12.

Finally, sample size places another limit on the distribution. In any small sample the likelihood of observing extreme values is small, despite these values existing in the general population. Therefore, on average a normative study based on a small sample will underestimate the true variability in the population. Somewhat counter-intuitively, the other effect of a small sample size can be to *over*emphasise differences in the general population, in the (statistically unlikely) event that a small number of individuals with extreme scores should appear in the sample. To state it simply, if a 'one-in-a-million' individual was included a sample the size of the current study, they would be assumed to represent a full 1% of the population, and so skew results markedly. Larger sample sizes reduce the effect of such outliers and are therefore clearly ideal in normative studies.

Summary of the Normative Data Computation Procedure

The procedure described represents the best attempt to match observed normative data to standardised scores. The issue of whether such a procedure is appropriate for particular variables has not been fully examined here, although a number of limitations to the procedure are described. While the use of standard scores permits comparisons between measures, this is clearly only of use if these comparisons are demonstrably valid. The specific application of these procedures to the research measures in this programme is examined in Chapter 8, *Results,* and Chapter 9, *Discussion.*

APPENDIX K. BOXPLOTS OF STANDARDISED SCORE GROUP PERFORMANCES.

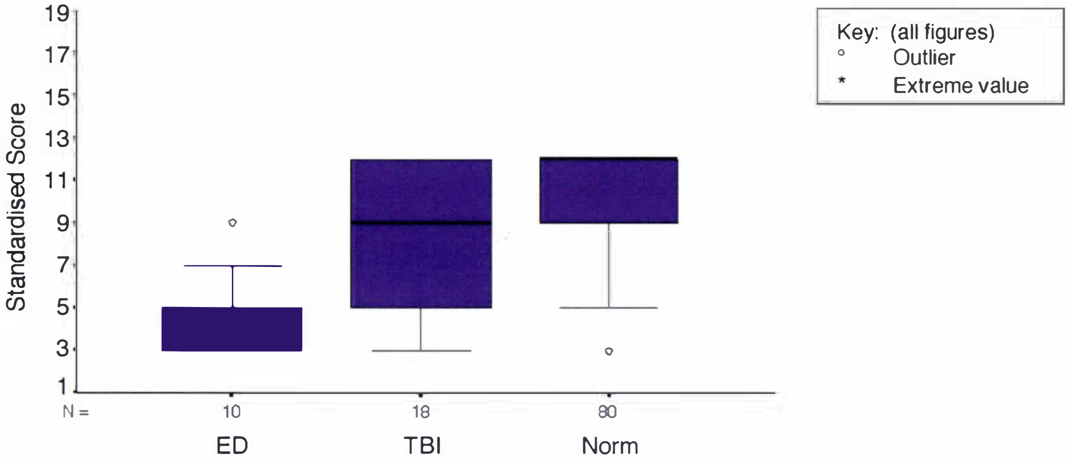


Figure A2. Boxplot of Graded Attentional Test standardised scores, by group.

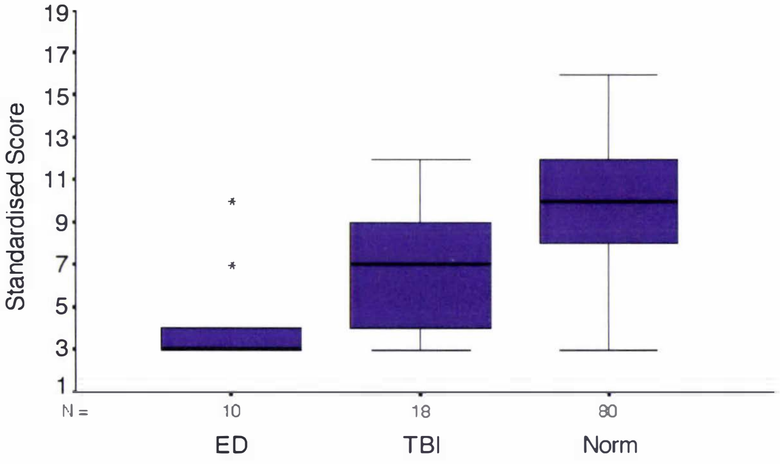


Figure A3. Boxplot of Digits Forwards standardised scores, by group.

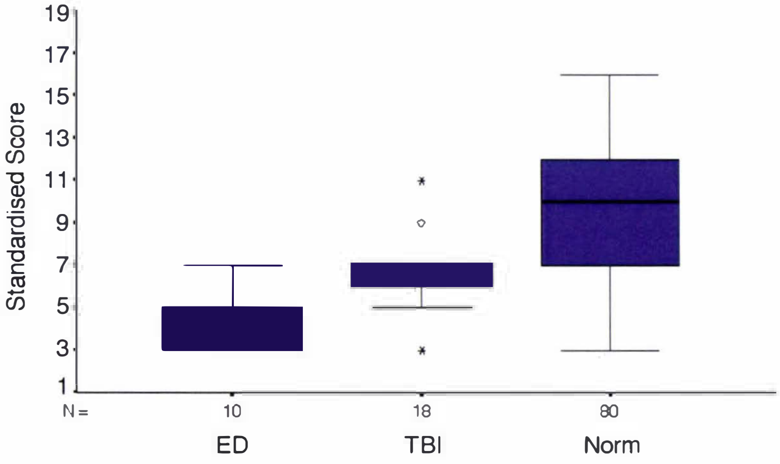


Figure A4. Boxplot of Digits Backwards standardised scores, by group.

Key: (all figures)
° Outlier
* Extreme value

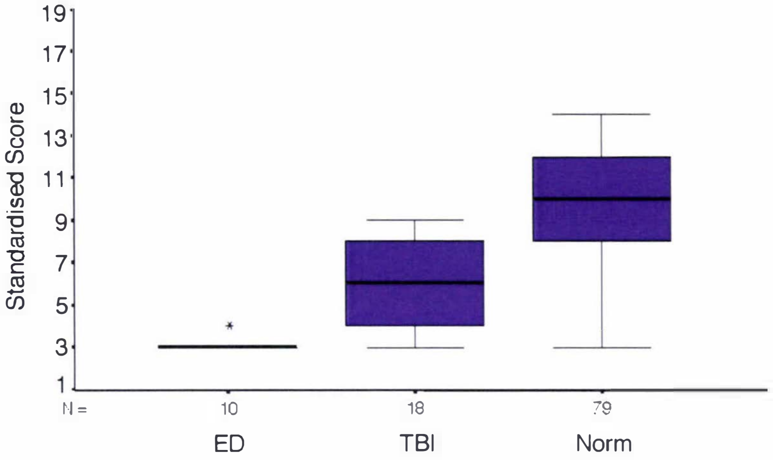


Figure A5. Boxplot of Verbal Paired Associates A standardised scores, by group.

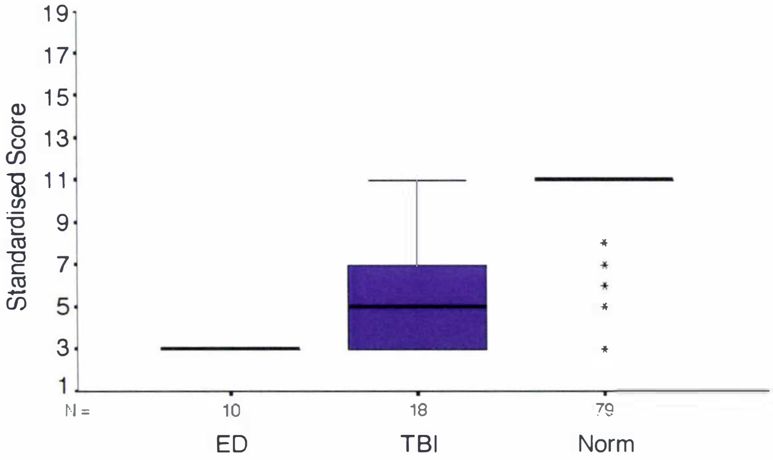


Figure A6. Boxplot of Verbal Paired Associates B standardised scores, by group.

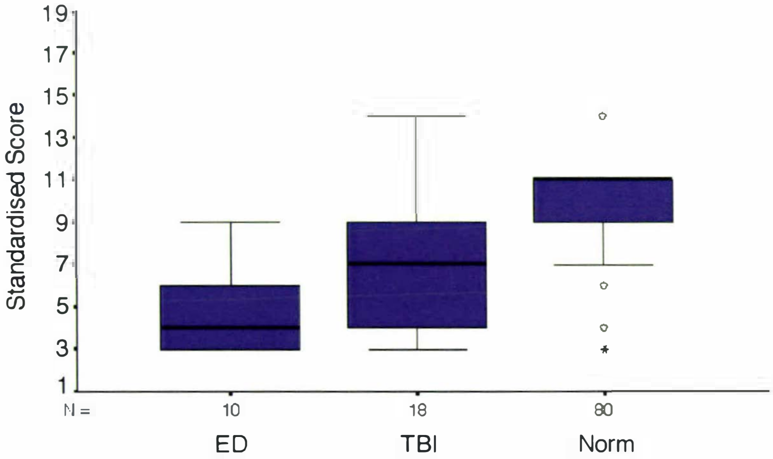


Figure A7. Boxplot of Logical Memory A standardised scores, by group.

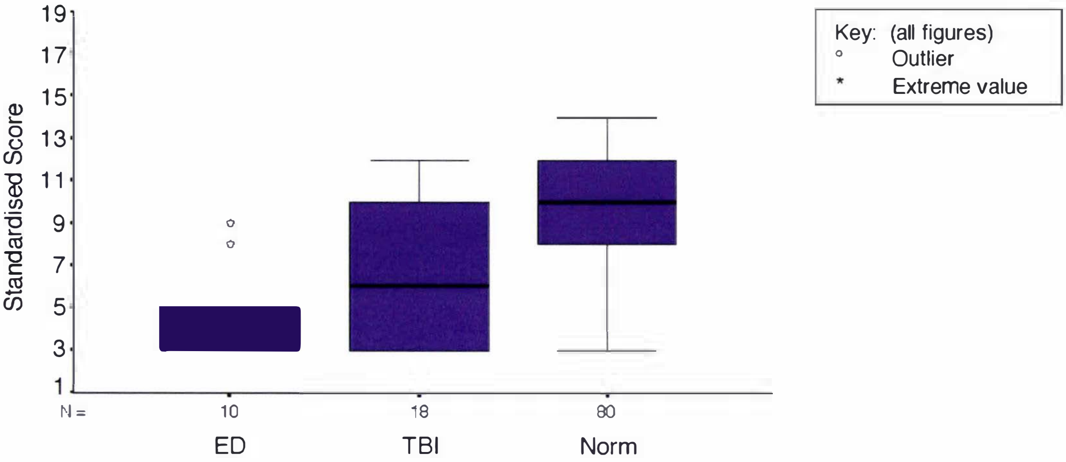


Figure A8. Boxplot of Logical Memory B standardised scores, by group.

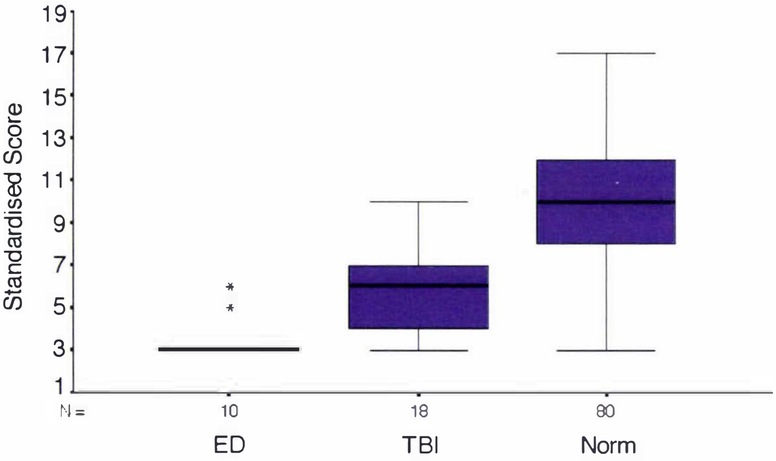


Figure A9. Boxplot of Faces A standardised scores, by group.

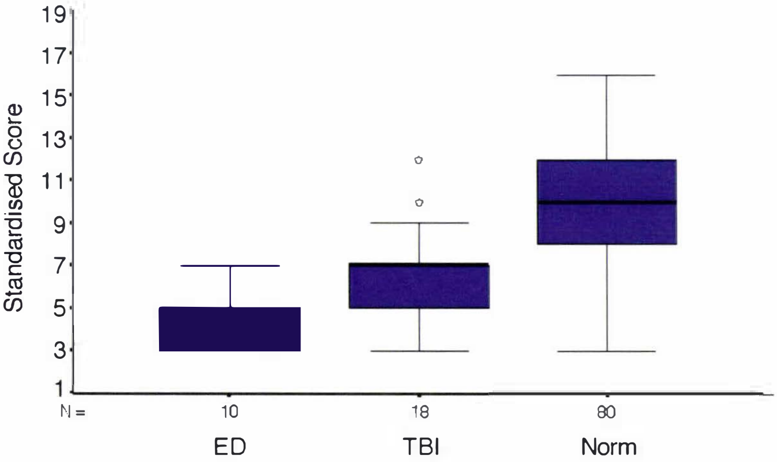


Figure A10. Boxplot of Faces B standardised scores, by group.

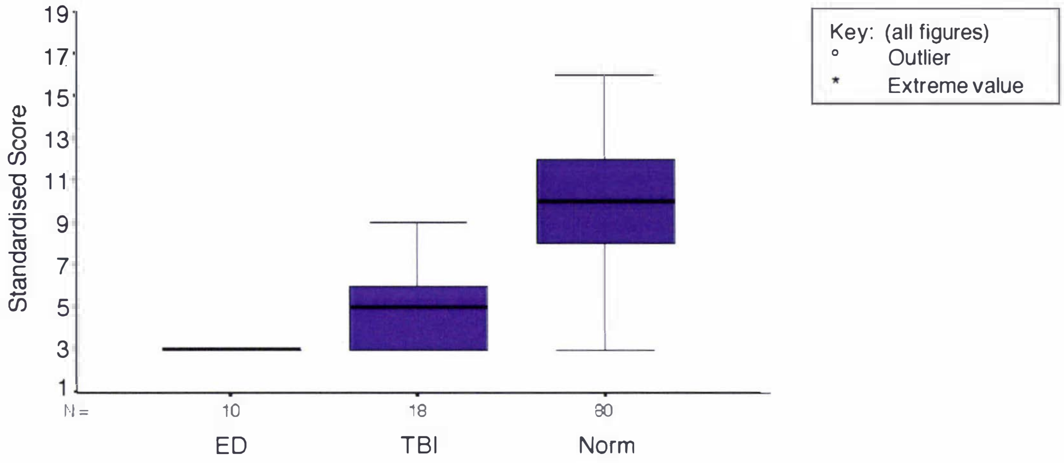


Figure A11. Boxplot of Family Pictures A standardised scores, by group.

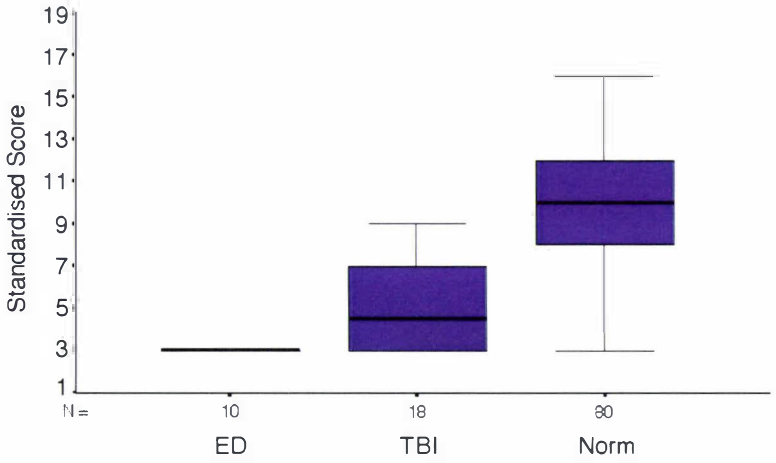


Figure A12. Boxplot of Family Pictures B standardised scores, by group.

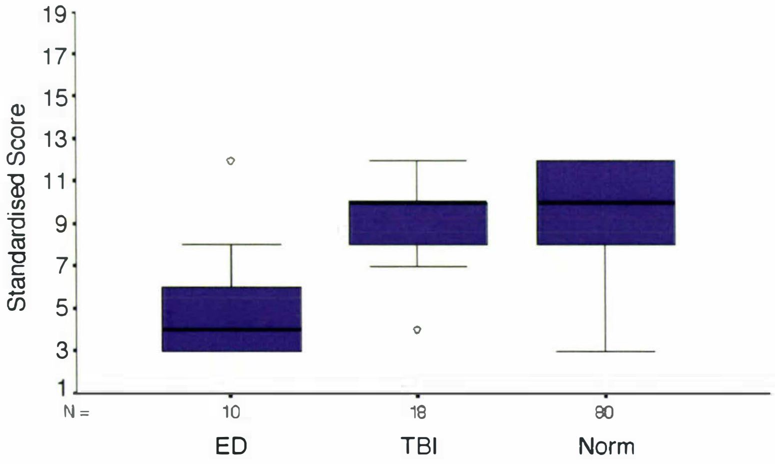


Figure A13. Boxplot of Auditory Reception standardised scores, by group.

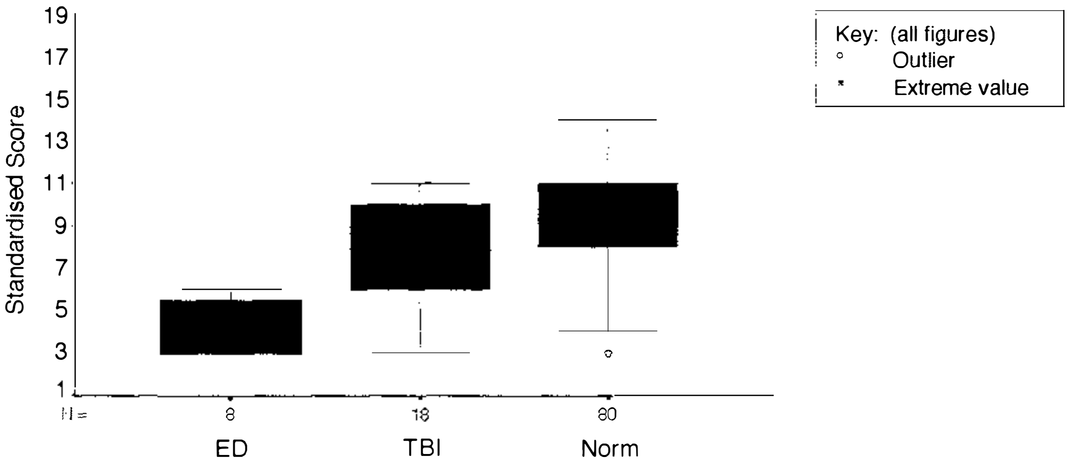


Figure A14. Boxplot of Boston Naming Test standardised scores, by group.

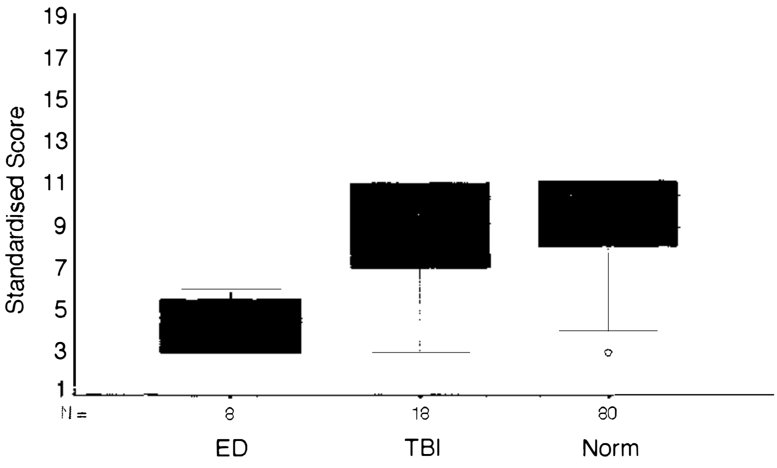


Figure A15. Boxplot of Boston Naming Test (all cues) standardised scores, by group.

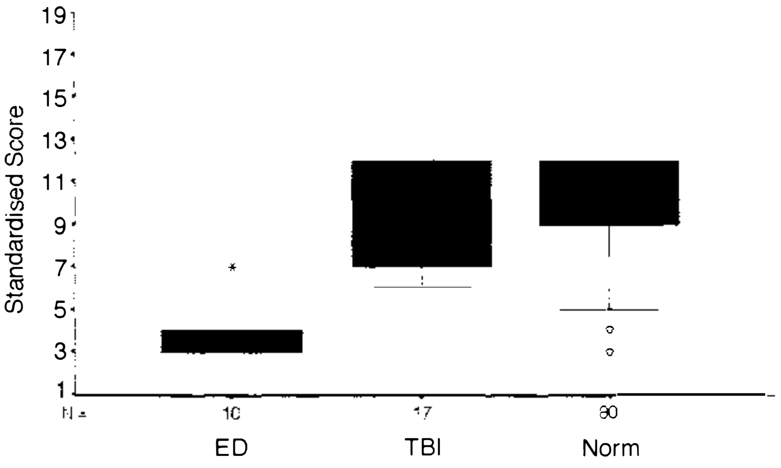


Figure A16. Boxplot of Familiar and Novel Language Comprehension Test Novel Language standardised scores, by group.

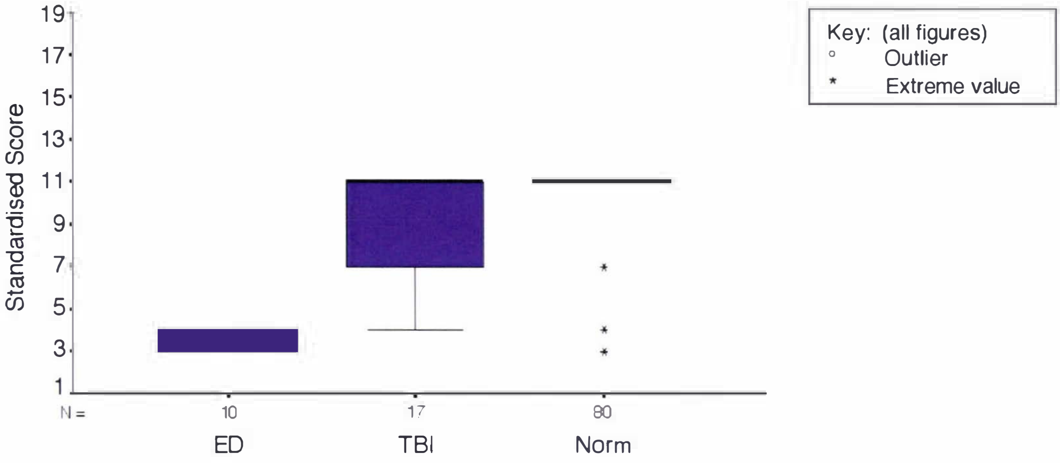


Figure A17. Boxplot of Familiar and Novel Language Comprehension Test Familiar Language standardised scores, by group.

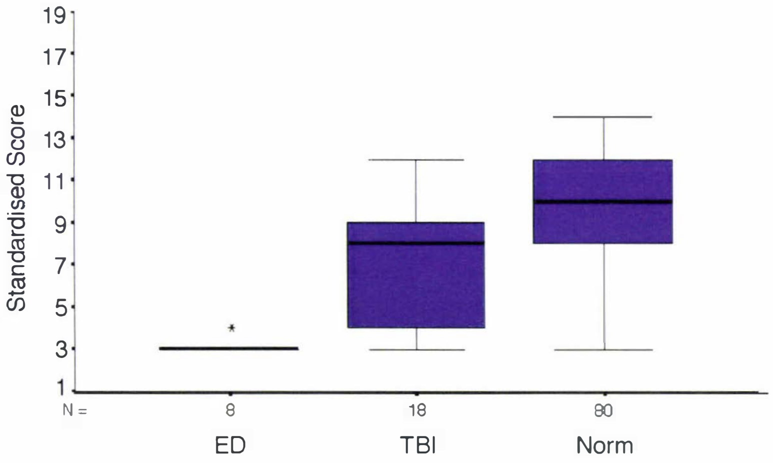


Figure A18. Boxplot of Hooper Visual Organization Test standardised scores, by group.

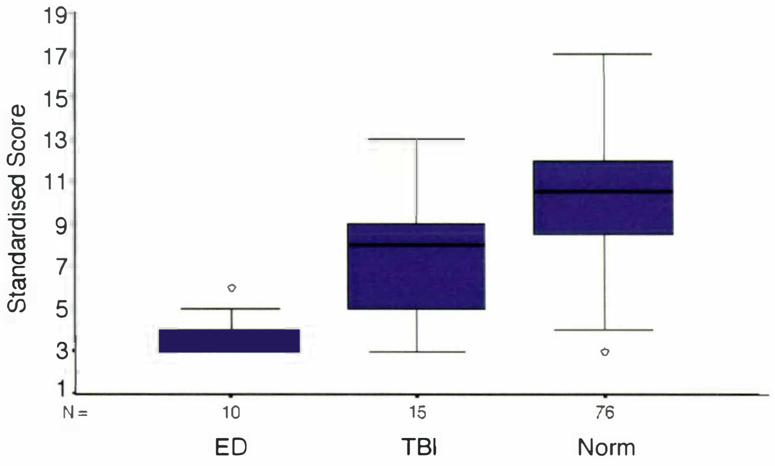


Figure A19. Boxplot of Matrix Reasoning standardised scores, by group.

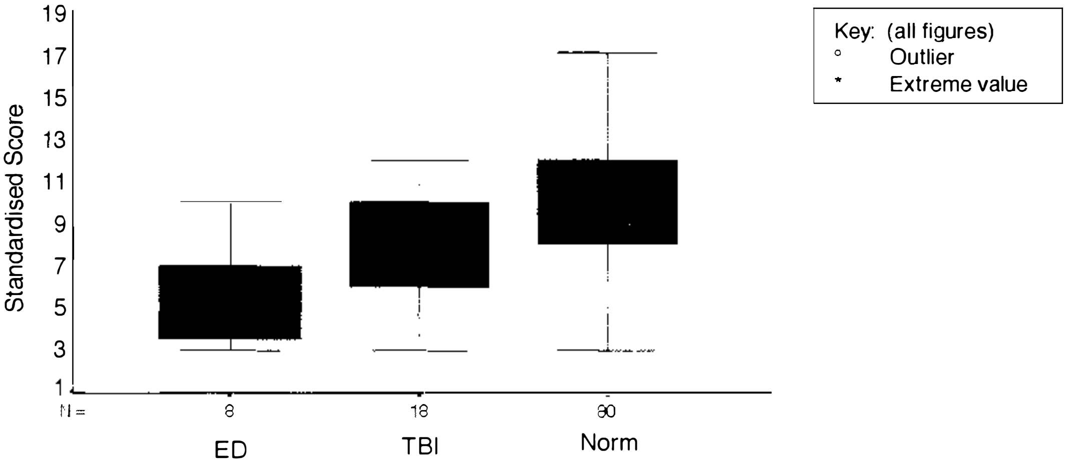


Figure A20. Boxplot of Match and Shift Categories Test Total Correct standardised scores, by group.

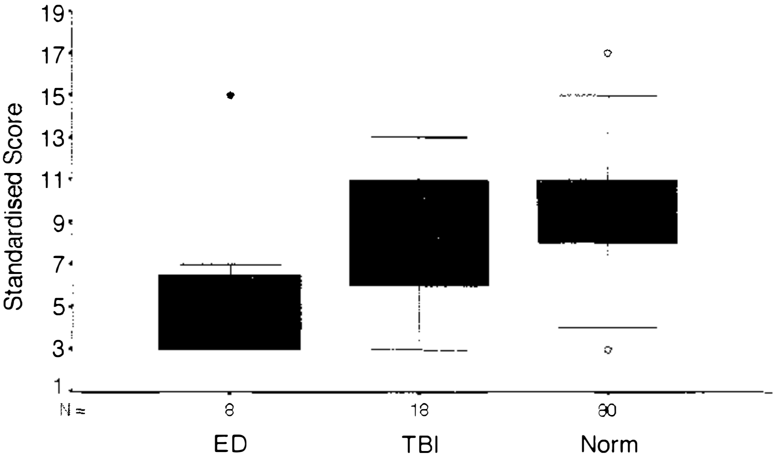


Figure A21. Boxplot of Match and Shift Categories Test Categories standardised scores, by group.

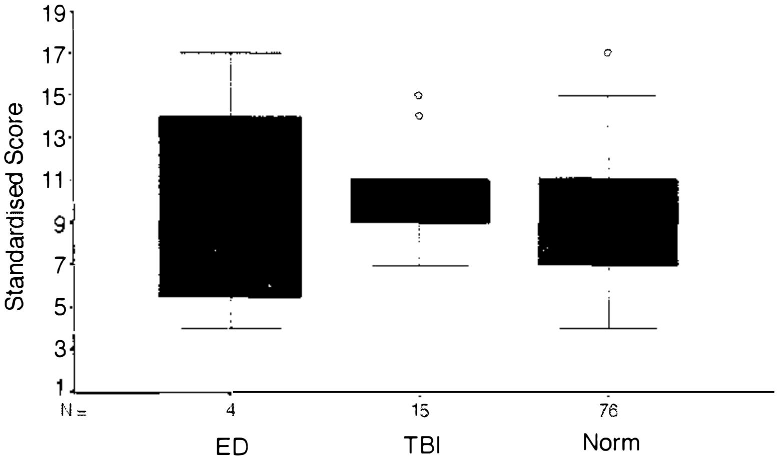


Figure A22. Boxplot of the Match and Shift Categories Test Perseveration standardised scores, by group.

It was a worthy challenge.