THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato
Research Commons

http://waikato.researchgateway.ac.nz/

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

THE DEVELOPMENT OF A

NEW ZEALAND ADULT READING TEST

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Social Sciences in Psychology

at the

University of Waikato

by

TRACEY JAYE HALLIDAY

University of Waikato

2006

ABSTRACT

The National Adult Reading Test (NART), developed in Britain is commonly used in clinical settings to estimate premorbid intelligence in New Zealand. Research suggests psychometric tests are more accurate if normed on the population they are used with. This study attempted to establish norms for the original NART based on a New Zealand population and develop a National Adult Reading Test for use with a New Zealand population (NZART). Sixty-four university students were administered the Wechslers Abbreviated Scale of Intelligence (WASI), the NART and the New Zealand Adult Reading Test (NZART). A regression equation was developed to estimate premorbid intelligence in this sample. Results indicate fewer errors occur on the NZART than the NART suggesting it may be a better indicator of premorbid intelligence for a New Zealand sample. Furthermore, the NZART was more accurate at estimating premorbid WASI IQ across all three subscales of the WASI at a range of IQ levels. Analyses were also conducted to ascertain the impact of demographic variables. Little overall difference was found in test scores in relation to gender, age or income. Although future studies need to be conducted to validate this new measure, initial results suggest that the NZART may be a more accurate predictor of premorbid IQ in a New Zealand population.

ACKNOWLEDGMENTS

I would like to begin by thanking all the people who made it possible for this thesis to be completed. I am grateful for having such wonderful supervision, without the support and encouragement of Nicola Starkey and Jo Thakker, this thesis would never have been completed. Thank you both for your patience and advice.

I would also like to acknowledge my participants and colleagues for having the courage to sit an intelligence test even though some of you were worried about the results.

I also owe a great debt to my husband Te Rongopai, (my rock), who has put up with so very much, worked long hours to support me, and who always believed in me. ‘Thanks babe’.

Thanks also to my children, Harlem, Cheyene and Okeroa who understood when I couldn’t be there, put up with my moods and helped out by doing all the chores I couldn’t get to, ‘big ups guys’. I also have to thank Roxanne and Kingi for taking care of the most precious of taonga when I needed a sitter, and also, thanks to Casey and Wiremu for their aroha and support, ‘Ka mau te wehi’.

I would also like to thank my sisters Kym and Lydia for their continual encouragement and faith in me that I could do this, and thanks to Rangi and Nicky and their respective families for being there after school when I wasn’t.

Lastly I would like to thank my parents, my family and my friends who each in their own way contributed to my being able to complete a thesis.

The year 2006 brought a great deal of joy and hardship for many of us, many tears were shed in both happiness and heartache. For those of you who have finished your journey, congratulations and good luck for the future, for those of you who are still in the trenches, ‘kia kaha’, you’ll get there.

Thank you for your support. I owe you all much love.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

Mankind has always been fascinated with the idea of testing their prowess and competence and the origins of testing date back centuries. Over 2000 years ago Chinese officials were taking tests of competence (Bowman, 1989) with components that assessed personality and mental ability (Liangshi, 1999) and history is littered with similar examples. For example, the ancient Greeks were known to test physical and educational ability (Doyle, 1974) and from as long ago as the Middle Ages, European universities relied on formal tests in awarding degrees and honours (Anastasi & Urbina, 1997).

A sense of moral responsibility brought about a rapid change in the type of testing during the 19th century. The need for uniform criteria in classification became apparent due to the inhumane treatment of the mentally ill (Anastasi & Urbina, 1997). The United States and European institutes worked towards establishing an objective system for identifying and classifying the mentally ill, and differentiating between the insane and the intellectually disabled. However, the early experimental psychologists of the 19th century were primarily concerned with establishing generalised descriptions of human behaviour rather than measuring individual differences (Anastasi & Urbina, 1997).

The development of testing changed dramatically with the well-known biologist Sir Frances Galton, a leading pioneer in psychological testing; who was primarily responsible for the development of the rating scale and questionnaire methods. His work led to early experiments with mental tests and the rise of intelligence testing in the late 1800's (Anastasi & Urbina, 1997). Cattell in 1890 described a series of mental tests administered to college students in order to determine their intelligence level. Later, after many years of hard work, Binet and his colleagues developed the Binet Intelligence scales. Originally developed to diagnose and classify mentally disabled children, the Standford-Binet was the first to use intelligence quotient (IQ) as a measure linking mental and chronological age (Anastasi & Urbina, 1997).

Later developments led to a number of intelligence tests, however it was soon recognised that these tests were limited in what they measured and the development of aptitude, achievement and personality tests began, in an ongoing search to more adequately assess human behaviour (Lezak, 1995).

The rapid evolution of neuropsychology began in the early part of the 20th century. In the 1940’s, heavily influenced by the work of scientists such as Broca (1865) and Warnicke (1874), neurology and psychology blended to develop its’ own identity (Cubelli, 2005). Initially born out of an escalating demand to accurately assess brain damaged and behaviourally disturbed individuals during wartime, neuropsychology became essential in screening, diagnosis and rehabilitation (Lezak, 1995). Clinical Neuropsychology evolved from this. It is an applied science primarily concerned with the relationship between the brain and behaviour. Clinical Neuropsychology focuses on the identification, assessment and treatment of individuals with brain dysfunction and testing is necessary in order to establish the degree of deficit, the origin and the location of the damage ((Eubanks, 1997; Sadock & Sadock, 2003).

A number of methods have been employed in the diagnosis of such brain damage; the usual clinical approach focuses on extensively studying the strength, efficiency and appropriateness of patients’ responses to certain questions (interviews, questionnaires and standardised tests) that can provide a relatively good indication of brain function (Lezak, 1995). In neuropsychology the assessment and diagnosis of brain dysfunction has centred mostly on cognitive functioning (which is the information handling aspect of the brain). This is mostly because a deficit in brain function will almost always result in some degree of cognitive deficit and also because deficits in cognition are easily measured.

Early investigators believed cognitive functioning could be attributed to one single factor - intelligence. Later developments suggested cognition is more complicated and involves a number of multifaceted systems (Lezak, 1995). These factors can be discretely and individually defined, however they are invariably intertwined and together produce a multitude of complex activities (Lezak, 1995). Identifying and examining one discrete function of cognition therefore requires the development and use of specialised tests. The primary aim of such tests is to achieve reliable, empirical and reproducible results that can be compared to tests of normal people of a similar age and demographic background in order to ascertain if and where there is a deficit (Sadock & Sadock, 2003).

In some patients a cognitive deficit is blatantly obvious and little testing is necessary. In others the loss is subtle and only apparent in complex situations. Other patients' deficits are so subtle that they are barely evident and only manifest as a very slight change in behaviour or emotion (Willishire, Kinsella & Prior, 1991; Franzen, Burgess & Smith-seemiller, 1997; Lezak, 1995). These latter forms of decline require more extensive testing. The easiest way to establish if an individual has suffered a decline in cognitive functioning is to compare his/her current performance to a pre-existing cognitive measure. However the existence of such measures is rare (Crawford, Venneri & O'Carroll, 1998). In these cases where the effects are extremely subtle it becomes the task of clinicians to determine the existence and extent of cognitive deterioration. As already mentioned this can be extremely difficult and, for this reason researchers have attempted to develop tests that produce a measure of premorbid IQ.

1. Measures of premorbid intelligence

The idea that a premorbid intelligence score can be obtained suggests that there is some prior level of cognitive functioning available and that it can be compared to an individual's current level of cognitive performance (Lezak, 1995). This level of comparison could be either individual or normed on a representative sample. However, given the already stated rarity of obtaining pre-existing individual scores; normative comparisons are the most commonly used. These are usually an average or median score of performance drawn from a well-defined population using variables such as age, gender and educational achievement (Lezak, 1995).

Over the years a number of methods have been employed for estimating premorbid intelligence, and these are described below. In order to accurately estimate premorbid IQ the test must fulfil three criteria. First the test must have adequate reliability, second it must be highly correlated with general IQ in the normal population and third the test must be resistant to the effects of neurological decline and psychological illness (Crawford, 1989).

1.1. Demographic variables

One of the most widely used means of estimating premorbid IQ is based on demographic variables. Demographic variables such as age, gender, social class and education level are highly correlated with an individual's intelligence test scores. Further because demographic variables are used, an individual's current cerebral dysfunction has no effect on the scores. The first method to be discussed here utilises a regression equation to predict IQ scores calculated from demographic variables and achieves two of the three criteria outlined by Crawford (1989).

Wilson, Rosenbaum, Brown, Rourke, Whitman and Grisell, (1978) developed a stepwise equation using age, sex, race, occupation and education to predict Wechslers IQ scores using the Wechsler Adult Intelligence Scale (WAIS). Their equation predicted 36% in Full Scale IQ (FSIQ), 53% of variance in Verbal IQ (VIQ), and 24% in Performance IQ. Education proved to be the best single predictor of all WAIS scales. Later Barona, Reynolds and Chastain (1984) utilised a similar approach to estimate IQ scores based on the Wechsler Adult Intelligence Scale-Revised (WAIS-R). The researchers added a number of extra demographic variables including rural/urban residence and geographic location to their formula. Dominant handedness was also included but later withdrawn as no correlation was found between this variable and IQ scores. Their equation predicted 36% in WAIS-R FSIQ, 38% of variance in the VIQ, 24% and in PIQ indicating demographic variables could be used as a reasonable estimator of premorbid IQ.

Initial studies seemed credible, with research suggesting that by using demographic variables a reasonable estimate of IQ could be obtained (Crawford, Stewart, Cochrane, Foulds, Besson & Parker, 1989; Stebbins, Wilson, Gilley, Bernard & Fox, 1990; Crawford & Allan, 1997). However other researchers showed disappointing results. Eppinger, Craig, Adams and Parsons (1987) found Barona's formula generally overestimated IQ in normal subjects. Further, the researchers noted some practical limitations with Barona's equation. For example they noted the occupational classification system failed to differentiate between those who had completed a masters degree and those who had obtained a bachelors degree. Another study using Barona's formula found low correlations between predicted and obtained IQ in a normal sample. Furthermore, these researchers noted Barona's equation failed to discriminate between brain damaged patients and normal subjects in their study (Klesges, Fisher, Vasey, & Pheley, 1985). Other researchers have criticised Barona's formula for generating a large standard error of the estimate for WAIS-R FSIQ (Silverstein, 1987; Stebbins et al., 1990). In addition, in a more recent study Griffin, Mindt, Rankin, Ritchie and Scott (2002) found Barona's method (which both over and under estimated FSIQ) to be inferior to other methods. Given the importance of developing an accurate method of measuring premorbid IQ other researchers have attempted to generate other possible solutions.

1.2. Present ability measures

An individual's current performance or present cognitive ability has been used to assess premorbid IQ. This method is based on two key assumptions: firstly that the individual's performance on one cognitive measure will infer performance on another measure, and secondly that not all cognitive measures will be equally affected by brain impairment Lezak, (1995).

One of the oldest methods to estimate premorbid IQ is based on present ability measures. It is known as the 'hold-don't-hold' method. Research suggests that some measures will be resistant to cerebral dysfunction and retain or 'hold' their ability to measure cognitive performance while others will be severely affected and will not ‘hold’ when individuals suffer from cognitive deterioration (Crawford, 1992). The first assumption has support; some cognitive measures do correlate positively in healthy individuals, for example vocabulary correlates highly with general intelligence (Crawford, 1992; O’Carroll, 1987). The second assumption has also been shown to have support; research shows some cognitive measures such as verbal ability are less affected by cerebral dysfunction than others such as arthimetic (Lezak, 1995).

The hold-don’t-hold method outlined above estimates premorbid ability based on an individual’s current performance on a measure that is considered to be relatively resistant to neurological impairment. Wechsler (1958) suggested hold tests; Vocabulary, Information, Object Assembly and Picture Completion were only minimally affected by age or brain impairment and thus they held their capacity to be useful as a measure of intellectual functioning. He recommended using a Deterioration Quotient that involved subtracting the age grade scores for the don’t-hold tests (Block design, Digit span and Similarities) from the hold tests. Unfortunately the WAIS deterioration quotient failed to be an adequate predicator of brain impairment (Lezak, 1995). In response to this other researchers suggested using an average of the scores obtained from the Vocabulary and Similarities subtests (McFie, 1975) or using only the score obtained from the Vocabulary test (Yates, 1956).

The Vocabulary subtest of WAIS has been one of the most widely used to estimate premorbid intelligence (Lezak, 1995). Initial research suggested verbal ability is retained in individuals across a wide range of brain disorders long after other functions such as ability in arithmetic have deteriorated. However the resistance of vocabulary to cerebral dysfunction has been questioned. Research shows that vocabulary is not as resistant to the effects of cerebral dysfunction as was commonly assumed (Crawford, Parker & Besson, 1988; Vandeploeg, 1994). Tests such as the WAIS vocabulary test require a verbal definition of a word, a function that tends to erode more quickly with brain damage than reading of a word or two (Lezak 1995). Crawford (1989) outlines a number of studies that reported vocabulary performance as significantly lower in a clinical population when compared to an unimpaired sample. Another more recent study also suggests the retrieval of verbal information may be significantly affected by semantic memory failure in patients with brain dysfunction (Patterson, Graham & Hodges, 1994).

In order to provide a solution to these issues, Lezak (1995) proposed the best performance method. This uses the highest level of current or past cognitive ability as a standard for comparison. The exception to this being when the highest score obtained comes from digit symbol, digit span or object assembly. These scales have a low intercorrelation with many other subtests and thus will be a poor estimator of premorbid ability. However not all researchers agree that the best performance method achieves adequate results. One study (Mortensen, Glade & Reinisch, 1991) found the best performance method over estimated IQ in a normal population. Participants were administered a battery of Wechsler scales. The results showed the highest test score was always higher than the IQ score because the IQ score is a mean score.

Another method outlined by Franzen et al., (1997) compares current scores with a large normative sample. If an individual's score falls below this norm they are labelled impaired. Lezak (1995) claims the results from this type of testing can be misleading and that it is only appropriate to directly compare an individual's score to a population norm if performance is not related to age, sex or education and the score is uniformly present in all individuals; factors which are almost impossible to achieve. The major drawback with this method is that an individual with an extremely high or low premorbid intelligence level may be disadvantaged. An individual with a high premorbid score may suffer a level of deterioration that places them above the cut-off point leaving the individual undetected. The reverse will be true of an individual with a low premorbid score; they may be labelled impaired without justification.

1.3. Single word reading tests

The idea that reading is related to general intellect led to a theory that by using single word reading tests an accurate estimate of IQ could be obtained. The final approach to be discussed here is one that estimates premorbid intelligence using single word reading scores. The first popular test of this kind was the Schonell Graded Word Reading Test (SGWRT) developed in 1942. In 1975, Nelson and McKenna conducted an initial study with normal subjects (*n*=98) and subjects with dementia (*n*=45). Using the SGWRT they demonstrated that word reading ability was highly correlated with general intelligence. Further they noted that word reading ability was well maintained in patients with dementia and that reading ability was more resistant to the effects of cerebral dysfunction than the WAIS vocabulary subtests (Nelson & McKenna, 1975). A later study confirmed these findings. Ruddle and Bradshaw (1982) administered the SGWRT to 78 non-impaired individuals and 75 patients with heterogeneous cortical diseases. Regression equations developed for WAIS VIQ, FSIQ and Performance IQ (PIQ) showed a reasonably accurate prediction of premorbid intelligence. However, although the SGWRT provided a reasonable estimate of premorbid intelligence in adults it was originally designed to assess children's reading ability and is subject to a ceiling effect when used with adults with above average intelligence (Franzen et al., 1997). Adults scoring 100 percent correct in the SGWRT have an equivalent IQ score of only 115 (Nelson & O'Connell, 1978).

1.3.1. The National Adult Reading Test (NART)

In response to their original findings with the SGWRT Nelson (1982) developed the New Adult Reading Test later renamed the National Adult Reading Test (NART). This has since become the most widely used test to estimate premorbid intelligence (Crawford, Allan, Cochrane & Parker, 1990; Crawford, Deary, Starr & Whalley, 2001). The NART, developed in Britain, consists of a written list of 50 irregular words presented in increasing difficulty that a subject is asked to read aloud. The words are relatively short to minimise the possible adverse effects of stimulus complexity that may occur in subjects with dementia. A NART error score is inserted into a regression equation to predict a WAIS FSIQ score. Verbal and Performance IQ scores can also be predicted using alternative equations.

Nelson and O'Connell reasoned that the average adult could decode normal English words even though they may not recognise them or know their meaning. However words that do not follow the normal grapheme-phoneme and/or stress rules (such as naive) cannot be decoded and an individual would need to be familiar with the words in order to pronounce them correctly, (Nelson & Willison, 1991). Their test construction relied upon an individual's premorbid recognition of words, rather than any current cognitive functioning. Using a list of such irregular words Nelson and O'Connell (1978) standardised the NART on a group of non-impaired subjects. The subjects (n=120) were presented with the NART, a short form of the WAIS and the SGWRT. Prediction equations and standard errors of estimates were developed for the WAIS VIQ, PIQ and FSIQ. Results showed the NART predicted 60%, 32% and 55% of the variance in WAIS VIQ, PIQ and FSIQ respectively. Furthermore, results also showed age and social class were not relevant variables in the relationship between reading level and general intelligence (Nelson, 1978). This is consistent with later research by Crawford, Stewart, Garthwaite, Parker and Besson (1988). These researchers tested the relationship between demographic variables and the NART. The results showed age and gender had no effect on obtained scores. However in contrast, both education (r=0.51, p<0.001) and social class (r=0.36, p<0.001) were correlated with NART scores.

In a later study, Nelson attempted to validate the original findings. A group of patients with bilateral cortical atrophy (n=40) were given the NART, a short form of the WAIS and the SGWRT. Their results were compared to the original standardised group. Results showed the patients were not impaired on the NART when compared to the control group but obtained significantly lower scores on the WAIS FSIQ, PIQ and VIQ further indicating the NART as an effective tool for predicting premorbid IQ. (Nelson, 1978).

A further cross validation by Crawford, Parker, Stewart, Besson and De Lacy (1989) reported the NART predicted 66% of the variance in VIQ, 72% in VIQ and 33% in PIQ, further validating the NART as a useful measure in estimating premorbid intelligence. Furthermore, the researchers reported no ceiling or floor effects despite having a wide IQ range in their sample.

More recently, the NART has been restandardised for use with the WAIS-R. Nelson and Willison (1991) presented the NART and the WAIS-R to 182 non-impaired participants. The results showed the restandardised NART to be a good predictor of premorbid IQ. Further analysis replicated earlier findings; that is, there was no correlation between age, social class and IQ scores.

1.3.2. The Wechsler Test of Adult Reading (WTAR)

Given the success of the NART, another single word reading test, the WTAR, was developed in 2001. The WTAR is based on the same methodology as the NART and is comprised of 50 single words. It was co-normed with the Wechsler Adult Intelligence Scale –Third Edition (WAIS III) and the Wechsler Memory Scale – Third Edition (WMS III). The WTAR was normed on American and United Kingdom (UK) adults aged 16-89. Test developers suggest the WTAR has high internal consistency in both the United States (.90 - .97) and in the United Kingdom (.87 - .95). They also report high test-retest reliability (.90 - .94).

Extensive studies by test developers suggest the WTAR is a reliable measure of premorbid IQ. Researchers administered it to Alzheimer’s patients and matched control subjects with the WTAR and found little difference in scores. Similar findings were obtained in a group of Traumatic Brain Injury (TBI) patients and a matched control group. The researchers also found little difference in WTAR scores in individuals diagnosed with Schizophrenia, Alcoholism or Depression when matched with control groups.

Unfortunately there is little outside empirical evidence to validate the reliability of the WTAR. Validation tests conducted by an outside authority would enhance the status of the WTAR. Reviews conducted on the WTAR suggest that although the WTAR correlates moderately with other similar reading tests, the WTAR group mean was 11 standard scores lower than the American NART (Thompson, 2005). This reviewer also noted that studies have yet to be conducted on predictive validity based on prior records of cognitive ability such as academic or military records. Ward (2005) criticised the WTAR for a lack of information on item selection and content validity information claiming the information given is at best vague.

2. Psychometric properties of the NART: A review of the literature.

The NART has been quoted as being the most reliable test for estimating premorbid intelligence (Crawford, 1992). It has been shown to be superior to Barona's demographic regression method (Crawford, 1992; Griffin et al., 2002) and the Schonell Graded Word Reading Test (Nelson, 1982). Furthermore, it has been used with patients of varying ages, and with a wide variation of diagnoses with great success.

Research shows the NART has high split-half reliability (0.90-93) (Nelson 1982; Crawford, Stewart, Garthwaite, et al., 1988), indicating high internal consistency, high test-retest reliability across a 10-day period (0.98) and high inter-rater reliability (0.96-0.98) (Crawford, Parker, Stewart et al., 1989). For example, in a study by O'Carroll (1987), twelve patients were administered the NART for inter-rater reliability checks; ten clinicians found correlations ranged from 0.89 – 0.99. In another study the hypothesis that the NART was highly correlated with general IQ was tested. Crawford, Stewart Cochrane, Parker and Besson (1989) used a varimax rotated factor structure to show the NART demonstrated a loading on the general intelligence factor of .85, higher than most of the WAIS subtests. Thus confirming the NART is highly correlated with general intelligence. Sharpe and O'Carroll (1991) confirmed this finding in a more recent study where NART performance was highly correlated with intelligence in the general population (*r*=0.77).

The validity of the NART has been tested in a number of studies and has demonstrated itself as a relatively robust measure in assessing premorbid IQ in a wide range of disorders. For example, Nebes, Martin and Horn (1984) compared 20 healthy individuals with 20 Alzheimer's patients and found NART scores did not differ across the groups indicating relative insensitivity to dementia. In another study Hart, Smith and Swash (1986) assessed twenty patients with Dementia Alzheimer Type (DAT) and compared these to fifteen healthy volunteers using the WAIS, the SGWRT and the NART. Using the regression equation developed by Nelson and McKenna (1975) the researchers found the NART was the best indicator of premorbid IQ. In a more recent study by Sharpe and O'Carroll (1991), twenty subjects with dementia were compared to twenty volunteers using a battery of tests including the NART and a short form of the WAIS-R. Results showed no significant difference between the NART scores of subjects with dementia when compared to the volunteers. However patients performed significantly lower on the WAIS-R.

The presence of mental illness does not seem to alter the validity of the NART. The research by Crawford, Besson, Bremner, Ebmeier, Cochrane, and Kirkwood (1992) examined the validity of the NART in two groups with schizophrenia. Thirty-five residents in long stay wards and twenty-nine patients residing in the community were matched for age, sex and education with healthy participants. Both groups scored significantly lower on the WAIS than the control group. However their NART scores did not differ significantly suggesting the NART is a valid estimator of premorbid intelligence. Similar findings were obtained in a longitudinal study with thirty-four treatment resistant schizophrenic patients over a 6- month period. Results showed scores remained stable over time (Smith, Roberts, Brewer and Pantelis, 1998).

The validity of the NART as a measure of premorbid IQ has also been tested among depressed individuals. A group of 39 depressed patients were compared to a matched control group. The NART and vocabulary subtest of the WAIS were administered to both groups. Depressed patients scored significantly lower than the control group on the vocabulary subtest, but there was no significant difference between the groups in NART performance (Crawford, Besson, Parker, Sutherland, & Keen, 1987).

In an attempt to ascertain predictive validity, Moss and Dowd (1991) administered the NART and the WAIS-R to a 20-year-old man who suffered a severe closed head injury. The results estimated his premorbid IQ to be 93; this was then compared to his pre-existing WISC-R score of 88 obtained at age thirteen. Wechsler (1974) reported that on average an individuals' obtained WISC score would be 6 points lower than an obtained WAIS score. The researchers concluded that the NART produced a very accurate estimate of their subjects pre-injury IQ.

The NART has also been tested across different age groups. Ryan and Paolo (1992) developed a regression equation for an elderly sample using 85 unimpaired American participants. This was then cross-validated on a sample of 41 elderly participants and retested on a sample of 20 patients with brain damage. The researchers concluded the NART provided an accurate estimate of premorbid IQ in all three samples.

However not all studies have found the NART to estimate premorbid IQ accurately. Freeman, Godfrey, Harris and Partridge (2001) administered the NART to 65 patients with TBI and 27 orthopaedic participants in New Zealand. The researchers concluded that 30% of the TBI sample was impaired on the NART. However the results from the research may be questionable as the researchers used a regression equation derived from United Kingdom demographic variables to calculate their scores. Franzen et al., (1997) suggests that applying a demographic equation to people outside of the normed population is problematic, as the equation is likely to be affected by shrinkage in a new sample. Further, Riley and Simmonds (2003) question the utility of using a regression equation based on a neurologically intact sample to assess TBI patients. TBI subjects are not representative of a normal population and are more likely to have a dual diagnosis of drug and/or alcohol abuse and are also more likely to have a lower academic performance.

Other studies suggest that this type of reading ability does remain constant after injury or with progressive dementia. For example, one recent longitudinal study assessed a sample of 26 people with TBI. Participants were administered the NART within 12 months of their injury and reassessed 12 months later. NART performance had significantly improved on the second test, suggesting a NART given within 12 months of TBI risks underestimating premorbid IQ (Riley & Simmonds, 2003). Additional research suggests that a degree of deterioration in single word reading ability may occur as dementia becomes more severe. One longitudinal study of Alzheimer’s patients found evidence of deterioration in NART performance over a three-year period (Fromm, Holland, Nebes & Oakley, 1991). The 18 patients were administered the NART annually and compared to 20 elderly participants of a similar age and education. The dementia patients pronounced fewer words correctly than the control sample at each yearly test. In a more recent longitudinal study Cockburn, Keene, Hope and Smith (2000) found a decline in NART scores in Alzheimer’s patients over a four-year period. The researchers tested 78 elderly patients with dementia annually over a four-year period. They found NART scores declined over time with the progression of the disease. However O’Carroll, Baikie, and Whittick (1987) found no significant deterioration in 30 elderly patients with dementia over a 1-year period. These conflicting results seem to provide evidence that over time the severity of dementia may impact on NART scores. However Crawford (1992) points out that where NART performance is impaired the level of dementia is generally already established and cognitive decline is readily apparent.

Studies have also been conducted to ascertain if the severity of dementia affects test scores. Stebbins et al., (1990) separated 199 elderly patients with dementia into three sub-groups (mild, severe and very mild dementia). The participants were administered the NART, the Wechsler Memory Scale (WMS) and selected scales of the WAIS-R. Results showed that in subjects with very mild dementia the NART did not provide an adequate measure of premorbid IQ as patients' NART scores did not differ from the control group. In subjects with mild dementia, IQ was underestimated and with severe levels of dementia the NART grossly underestimated IQ. Taylor (1999) replicated these findings in a study where 58 Dementia Alzheimer's Type (DAT) patients and 58 patients with Multi-Infarct dementia (MID) were given a large number of tests including the NART. The researchers found the NART underestimated premorbid IQ and that NART performance was severely affected by the level of neurological impairment. Patterson et al., (1994) also found the NART score correlated with the severity of dementia. Their study of 45 patients with DAT of varying degrees showed a dramatic decrease in NART scores as a function of dementia severity. The researchers suggest the deterioration is a consequence of semantic memory failure. In contrast, O'Carroll and Gilleard (1986) examined thirty elderly patients with dementia to assess the sensitivity of the NART to dementia severity and found no significant differences between DAT patients and MID patients when compared. These studies indicate that in some cases the NART may be insensitive to dementia severity but not in others, however the reasons for these conflicting results have not been fully explored.

A further question has been raised as to the applicability of the NART over certain neurological conditions. Crawford, Parker & Besson (1988) examined the validity of the NART as a premorbid estimate of IQ in a number of organic conditions including DAT, MID, Huntington's disease, alcoholic dementia, Korsakoff's psychosis and closed head injury (CHI). Patients (*n*-70) were matched for age, education and sex with healthy individuals and presented with the NART and the vocabulary subtest of the WAIS. Results showed that NART performance was not significantly different for alcoholic dementia, DAT, MID and CHI patients when compared to control subjects. However, for Huntington's and Korsakoff patient's scores were significantly lower than the unimpaired individuals suggesting the NART may not be suitable for these two conditions. More recently O'Carroll (1992) compared scores of Korsakoff patients and healthy control subjects and concluded that the pathology affected NART scores. In another study by Ebmeier, Booker, Cull, Gregor, Goodwin and O'Carroll (1993) the validity of using the NART to estimate premorbid IQ in long-term survivors of hemispheric glioma and whole brain irridation was examined. Sixteen patients were matched for age, education and class with healthy participants. Results suggested that NART scores were highly correlated with current IQ but not with premorbid IQ.

Given the difficulty of assessing premorbid IQ and the overwhelming number of possible pathologies and manifestations of those pathologies it is easy to see how a wide range of conflicting results could emerge. Any number of factors including time, assessment procedures, severity or particular manifestation of pathology could alter the results of testing. These inconsistencies seem to create a paradox and it would seem to some extent to be naïve to suggest any measure could serve as a catchall. On the other hand given the importance of estimating premorbid IQ, a reasonable and reliable method must be used and the majority of researchers suggest the National Adult Reading Test is the best of such measures.

3. Modifications of the NART for other countries

Of all the many possible mitigating factors that could influence psychometric testing cultural variation has received little attention (Ardila, 1995; Lezak, 1995). Yet researchers agree that ethnicity and culture do affect test scores (Ardila, Rosseli & Puente, 1994). The NART was originally developed and normed on a British population. The scoring procedures are based on British pronunciation and British familiarity with words. For example words such as gaol or drachm may not be familiar to other ethnic groups (Franzen et al., 1997). Early research conducted with the original NART in countries outside of Britain have generated inconclusive results suggesting possible problems with using the test outside of the population it was normed on.

In reaction to this, Swartz and Saffran (1987) developed the American National Adult Reading Test (AMNART) standardised on 109 normal adults in North America. They replaced 23 words from the original NART with American words. In agreement with Nelson (1982) they found high correlations between AMNART predicted IQ and the WAIS IQ measures (r=. 72, .51, and .72 for VIQ, PIQ and FSIQ respectively). In a validation study Grober and Sliwinski (1991) administered the NART, the AMNART, and an abbreviated version of the WAIS to a group of normal adults (n=40). Thirty-five of the participants made fewer errors on the AMNART, suggesting the AMNART as an easier measure for the North American population. In a more recent validation study Boekamp, Strauss and Adams (1995) evaluated the validity of the AMNART with dementia patients and normal individuals from two different ethnic groups. The AMNART and WAIS were administered to 30 males with a mild to moderate degree of dementia and a control group of 30 males of comparable age and education. Results showed little difference between scores of White and African American participants. Overall African Americans scored slightly lower in both tests. The greatest difference was found between dementia patients and their matched subjects. Results showed patients with dementia performed substantially lower on the WAIS-R VIQ than the control group, but there was little difference in AMNART scores. The researchers concluded the AMNART worked equally well for both ethnicities as a premorbid measure of IQ.

Other studies showed similar success. In 1989, Blair and Spreen modified the NART for use in both America and Canada. They added 29 words, deleted 17 and ensured scoring was based on American and Canadian pronunciation. They presented their test, the New Adult Reading Test-Revised (NART-R) also know as the North American Adult Reading Test (NAART) and the WAIS-R to 17 Americans and 49 Canadians without history of brain dysfunction. Correlations of the NAART with FSIQ, VIQ and PIQ were .83, .75 and .40. The NAART also demonstrated an inter-rater reliability of .99, and internal consistency of .94. Scores were combined with demographic variables to generate multiple regression prediction equations for WAIS–R

IQs. The researchers found that the demographic variables (age, gender, occupation and education) were not significant in the prediction of IQ and concluded that their revised version of the NART was a reasonable predictor of premorbid IQ.

Additional research by Berry, Carpenter, Campbell, Schmitt, Helton and Lipke-Molby (1994) further supported Blair and Spreens' (1989) NAART as a premorbid measure of IQ. NAART scores were used to estimate WAIS-R IQ and were correlated with scores obtained 3.5 years earlier from 54 healthy adults. Participants were both male and female, white Americans with a mean age of 67.8 years. Obtained scores correlated reliably with previous findings; FSIQ .70, VIQ .68, and PIQ .61. The researchers concluded that their results confirmed the retrospective accuracy of the NAART in predicting WAIS-R scores in older subjects over time.

Other similar studies validated the NAART as an accurate estimator of premorbid IQ in a clinical sample. Johnstone, Callahan, Kapila and Bouman (1996) administered the NAART and WAIS-R to 232 neurologically impaired patients from different ethnic backgrounds. NAART and WAIS-R scores were comparable for all three subgroups (DAT, TBI and other neurological impairment). The researchers concluded that the NAART was a useful predictor for premorbid IQ across these three subgroups. In a more recent study Uttl (2002) administered the NAART to 351 healthy adults. Results indicated that on average males made more errors than females on the NAART but obtained higher WAIS-R scores, but overall it was concluded that gender had no significant effect on NAART scores. Further, NAART scores increased only slightly with age (4.5 points across lifespan) and with education (1.5 points per year of education). The research indicated the NAART was a reliable predictor of premorbid IQ for young, middle-aged and older adults.

However not all researcher have found positive results when using the NAART. Wiens, Bryan & Crossen (1993) found the North American revised test to be only a modest estimator of premorbid IQ in an unimpaired sample. A battery of tests including the WAIS–R and the revised test were administered to 302 individuals from different ethnic backgrounds. The sample comprised both male and female subjects and ages ranged from 20 – 54 years. The researchers found no age or gender effects and only a minimal effect from level of education (*r* = - .14 between education and NAART errors). However the researchers did note a difference in the males and females NAART error scores. Males made more errors on the NAART (27.2) than females (25.5). Correlations of the NAART with FSIQ, VIQ, and PIQ were slightly lower than previous findings .46, .56, and .22 respectively.

Other countries have also begun to develop reading tests that are more suitable for their own populations. Australia has adapted the NART for use in their country. The modification of the NART was developed using pronunciation given in an Australian dictionary rather than that provided in the British manual. Results indicated the modified NART was a valid measure for estimating premorbid IQ in an elderly sample with dementia (Willshire, Kinsella & Prior 1991). However earlier studies indicate the original British NART was effective in estimating premorbid IQ in an Australian population. The NART, SGWRT and the Wechsler Memory Test (WMS) were administered to 65 elderly volunteers and 16 patients with DAT. Results showed the WMS and NART combination was effective in assessing premorbid IQ. Further analysis reiterated what other researchers had pointed out; gender was not a relevant factor in predicting IQ scores (Schlosser & Ivison 1989).

Other countries such as France, Italy and the Netherlands have also developed specific versions of the NART that are more suitable for their own population (Isella, Villa, Forapani, Piamarta, Russo & Appollonio, 2005; Mackinnon & Mulligan, 2005; Schmand, Geerlings, Jonker & Lindeboom, 1998).

4. The rationale for developing New Zealand tests

The rationale for developing psychometric measures for individual countries is easy to justify. Researchers agree that individuals from different cultural backgrounds perform differently on cognitive tests (Helms, 1992). Anastasi & Urbina, (1997) suggest neuropsychological tests favour the culture the test was written for. In addition a number of researchers propose that the evaluation itself is only useful if the individual is compared to norms that can tell us about people in the general population who have similar demographics (Lezak, 1995; Anastasi & Urbina, 1997).

Diagnosis of cognitive dysfunction is difficult enough without having to interpret whether the individual's results are valid in relation to the appropriate norms for interpreting that individual's performance (Harvey & Siegert, 1999).

It is obvious that neuropsychological tests need to be developed for a wide range of different cultural contexts. For example dementia, one of the more prominent brain disorders, affects one in twenty people over 65. With an increase in age there is an increase in risk; one in five New Zealanders over 80 develop dementia (Harvey & Siegert, 1999). Furthermore, due to the wide variation in symptoms, dementia is one of the more difficult disorders to diagnose. For New Zealanders the task is made more difficult due to the reliance on psychometric tools normed in another country. Despite continued cautions in relation to validity (Ardila, 1995; Lezak, 1995; Franzen et al., 1997) clinicians here typically rely on overseas norms to interpret New Zealand tests. Another common cause of brain dysfunction is TBI (Kurtzke, 1984), which affects over 900 New Zealanders each year (Barnfield & Leathem, 1998). Like dementia, TBI is difficult to assess. The extent of dysfunction is dependant on a number of variables including age, severity, location of damage and how the injury occurred. Diagnosis can also be very difficult; severity can range from an individual with a slight personality change to a patient who is in a vegetative state. In order to accurately assess and diagnose these disorders, New Zealand clinicians again rely on overseas norms. The results of such tests may lead to an over identification of deficits that have more to do with the cultural bias of the test, than the dysfunction of the individual (Ogden, 2001; Barker-Collo, 2001; Ogden & McFarlane, 1997).

It may be argued that many brain pathologies seem comparable across cultures, however cultural learning influences performance on psychometric tests and thus an error in assessment will occur when using an assessment tool developed by another culture even when the dysfunction gives the impression of being similar (Ardila, 1995). The NART and dementia are good examples of this; although the symptoms of dysfunction may appear similar across cultures, the cultural influence of the NART may affect the testing process. As previously mentioned, some of the words used in the original NART are not words that are typically used in other countries. The words reflect a British population and other cultures may not be familiar with them (Swartz & Saffron, 1987; Blair & Spreen, 1989; Willshire et al., 1991; Franzen et al., 1997). Words such as cellist and assignate fail to appear at all in the New Zealand Oxford Dictionary (2005). Other words such as demesne, campanile and drachm are somewhat outdated and uncommon to a New Zealand population. This can then only lead to an individual being disadvantaged causing error in assessment that basically renders the test invalid. Establishing appropriate norms for a New Zealand population will increase the accuracy of diagnosis and assessment in this area.

The rationale for developing relevant tests is easy enough to justify, but the reality is that most psychometric tests used in New Zealand have been standardised on North American or U.K samples (Ardila, 1995; Lezak, 1995; Franzen et al., 1997), there is an overwhelming lack of norms for a New Zealand population and little empirical research has been completed to date.

5. New Zealand and psychometric tests

Although culture has been largely ignored in the development of psychometric tests, (Lezak, 1995) it has not gone completely unnoticed and is slowly gaining ground. Researchers within New Zealand, along with other countries are becoming more aware of the problems inherent with using psychometric tests developed by and for a North American population.

Research that has compared New Zealand and United States norms have found substantial performance differences. Heriot and Beale (1996) assessed children’s learning disability using the Visual Sequential Memory (VMS) subtest of the Illinois Test of Psycholinguistic Abilities, normed on a North American population. Two hundred and thirty eight children aged from six to 16 were administered the test. Results showed significant differences between the performance of American children and New Zealand children. Overall children from New Zealand performed better than American children, the largest difference being for six year olds who performed significantly better than the American group. The researchers suggested using the test with caution in a New Zealand sample and postulated different cultural backgrounds, different exposure to school and educational abilities as well as different life factors as possible reasons for variations in scores.

In another study Barnfield and Leathem (1998) assessed fifty TBI patients of Māori and European ethnicity with a number of cognitive tests developed overseas. Results showed Māori preformed at a significantly lower level than non-Māori in all tests. Researchers suggest cultural bias may have affected the testing. Similar results were found in another study examining words. Rolleston (2001) tested the validity of the State Trait Anger expression Inventory (STAXI) on a New Zealand population. Two studies were conducted, the first analysed existing STAXI scores from 197 male prison inmates. In the second study 159 male prisoners were administered a reworded STAXI. The two sets of data were then compared. The researcher found that New Zealanders had particular difficulty with some of the word items not typically used in New Zealand (e.g. pout). By rewording the STAXI with more familiar words the researcher found internal consistency improved; the reworded STAXI better reflected a New Zealand population than the original. It has been hypothesised that language is influenced by cultural and sub cultural backgrounds, and may be a particular problem where the language or vocabulary is subject to international differences (Harvey & Siegert, 1999).

Another hypothesis suggests that words themselves are an area for concern when testing. The problem arises when unfamiliar words from another culture are used in assessment. In a recent study examining naming deficits and word retrieval difficulties, the Boston Naming Test (BNT) was administered to 58 New Zealand university students from different ethnic backgrounds. Participants' ages ranged from 17 to 25, and both male and females were included in the sample. Barker-Collo (2001) noted New Zealand subjects scored significantly lower than American and Canadian subjects on the BNT (1.2 standard deviations below the American sample). This low score placed the New Zealand participants in the 10th percentile and would result in them being identified as impaired. Further analysis revealed that New Zealand and Australian populations scored significantly poorer in word items that were not frequently used in those countries, for example for words such as beaver the error difference was as great as 60%. Māori in particular made significantly more errors than European individuals. Cultural bias was concluded to be the leading mitigating factor to the results found. Factors such as age and education that had been previously associated with higher BNT scores were ruled out as possible mitigating variables. The New Zealand sample was both younger and had obtained a greater level of education before testing than the American sample, suggesting they should have scored higher than their American counterparts.

Barker-Collo, Clarkson, Cribb, Grogan (2002) also found similar results when they examined American word content and its affect on test scores. Fifty-six healthy New Zealanders aged 17 to 47, both male and female and from different ethnic backgrounds were administered the California Verbal Learning Test Performance (CVLT) and an alternative (NZ-VLT). Results showed New Zealanders scored poorly on the CVLT. Participant's scores placed the average New Zealander below the 16th percentile, suggesting a possibility that an over identification of deficits could occur. However performances were improved on all NZ-VLT trials. The researchers hypothesised that the content of a verbal test can affect the test performance and that culture is invariably related to content. They further suggested caution in using assessment measures with individuals who do not match the standardisation sample.

They further encourage the development of appropriate norms for different populations and cultural groups.

Another later study by Barker-Collo (2003) compared 137 New Zealand university students from different ethnic backgrounds to normative data in the United States. Participants were administered the Symptom Checklist-90-Revised (SCL-90-R) and the profile of Mood States (POMS). The results showed a significant difference between New Zealanders and Americans in regards to psychopathy. New Zealanders scored higher on obsessive compulsive, phobic and anxiety scales, but no significant difference was found between the two groups in relation to the mood. The researcher concluded that cultural identity was a major factor in the resulting difference.

One final study to be mentioned here was conducted in New Zealand and attempted to validate the NART as a useful measure of estimating premorbid IQ in an unimpaired sample (Fisher, 1996). The researcher administered the NART, the WAIS-R, the Becks Depression Inventory and the State-Trait Anxiety Inventory to 100 normal participants. Both male and female participants took part and came from different ethnic backgrounds, with a mean age of 31. Results showed the NART overestimated WAIS-R FSIQ at the lower end of the IQ range by 17 points and underestimated the WAIS-R FSIQ at the higher end by 18 points. Further analysis revealed gender affected NART scores. Females made fewer errors on the NART than males (23.11 and 28.39 respectively). However this difference was not significant. Furthermore, Fisher (1996), found age had no significant effect on obtained scores, but NART scores and education were correlated (r=-.38, p<0.001).

Given the recent awareness of using overseas tests on individuals with other backgrounds, and the move by other countries to develop their own standardised tests, it would seem feasible for New Zealand to also begin undertaking such a task. The results thus far in relation to the NART suggest the development of a New Zealand NART is a step in the right direction.

6. Aims of the current study

The current study has three aims. Given that previous studies have shown wide variation regarding the effects of demographic variables in assessing premorbid IQ the first aim is to ascertain whether demographic variables have a significant effect on the NART, WASI or NZART scores in a New Zealand population. It is hypothesised that demographics such as age, gender and income will have little effect, while education and ethnicity will have a significant effect.

The second aim of this study is to develop New Zealand norms for the NART and determine if these norms are more accurate than the original NART. It is hypothesised that the New Zealand normed NART will be superior in that it will be more accurate than the original NART.

The final aim of this study is to develop a New Zealand Adult Reading Test using New Zealand words and developed using a regression equation to better estimate premorbid IQ. It is hypothesised that the NZART will be more accurate than both the New Zealand normed NART and the original NART.

Method

1. Pilot study

Ethical approval was obtained from the Department of Psychology, University of Waikato Psychology Ethical Review Committee. Following this an initial pilot study was conducted to ascertain the suitability of words for the New Zealand Adult Reading Test (NZART). A large number of irregular words (*n*= 72) were selected from the New Zealand Oxford Dictionary (2005). The words were selected on the premise that they did not follow the normal grapheme-phoneme and/or stress rules, and that they were likely to be known to, or be encountered by New Zealanders (this was decided by discussions with participants, university staff and the scanning of contemporary literature).

These 72 words and the 50 original NART words were recorded in written form together with the correct pronunciation according to the International Phonetic Alphabet. (A full list of the words is included as Appendix A, page 70). This was then presented to a linguistics professor at Waikato University who recorded the words on to a dictaphone. This ensured the researcher understood the correct pronunciation and could accurately assess the performance of the participants.

Participants for the pilot study were selected from an opportunity sample on a random basis (*n*=20) from around the university campus. The participants were given a brief overview of the research in both written and oral form explaining confidentiality, the right to withdraw and how to obtain the results of the research, they were also asked to complete an informed consent form. Once this was completed the subjects were asked to read aloud into a dictaphone the 72 selected NZART words randomly combined with the 50 original NART words. They were instructed to read the words as best they could at a pace that they were comfortable with.

All answers were recorded on to a dictaphone and were later scored for incorrect and incorrect pronunciation. Participants' were asked about their familiarity with the words, how easy they found it to decode words they did not know and what words they found to be easy or hard to pronounce. Notes were made of participant's answers and these notes were used in the selection of a final word list. Words were deleted for a variety of reasons. For a full list of words and reasons for deletion see Table 1.

Some words such as capon, eucharist and apophthegm were removed because participants' found they could decode them even though they had never encountered them before. Others such as wyvern, gaoled and campanile were taken out as participants agreed that these were not words they had encountered before.

Table 1.
Reasons for deleting words

Reason	Words		
Not in the New Zealand Dictionary	Assignate Cellist		
Words with a Low Inter-rater Reliability	Prelate Aeon	Puerperal Sidereal	Aver
Words that could be Decoded	Radix Capon Eucharist	Synapse Epistle Banal	Apophthegm
Words decided by clinical psychology students and psychology professor to be ambiguous	Cheyenne Beatify	Ci-Devant Bourgeois	Marquess
Words unlikely to be encountered in New Zealand	Drachm Campanile	Gaoled Wyvern	Epergne

Five words, aeon, sidereal, prelate, aver and puerperal were deleted as earlier studies by Crawford et al, 1999 found that these words had very low inter-rater reliability. Two more words, cellist and assignate were deleted because they are not in the New Zealand Oxford Dictionary (2005).

Once this list was produced it was given to two clinical psychology students and a psychology professor. They were asked about their familiarity with the words. It was decided after this discussion that five more words should be deleted as they could be mispronounced due to New Zealand dialect or pronounced in more than one way these were bourgeois, beatify, cheyenne, marquess and ci-devant.

In order to create a final list of words a procedure based on Fromm, et al. (1991) was adopted. The number of words pronounced correctly by each participant was tabulated and arranged in order from highest number of correct to incorrect. Words were then chosen that spanned the entire range (e.g. words that everyone pronounced correctly to words that no-one pronounced correctly). Words were chosen randomly from each group of words. For example, if ten people pronounced six words correctly three of those words will have been included in the NZART.

2. Recruitment

Participants were recruited from the University of Waikato via flyers posted throughout University. Potential participants were offered a five-dollar gift voucher from the Warehouse in recognition of their time or if they were first year psychology students they were offered an alternative 1% credit towards their final grade for every hour they participated. Those that were interested were asked to make contact with the researcher via e-mail to arrange a suitable time and date for the assessment. A total of seventy participants volunteered to take part in the study. Six participants were excluded as English was not their first language and one participant was excluded due to very low scores in all test measures.

In order to ensure the respondents ability to complete the required tests without possible confounding variables the following criteria for inclusion in the study were employed a) English as a first language, b) No history of substance abuse, c) No history of psychological illness, d) No reading or eyesight problems that would affect their ability to undertake the test, e) No history of head trauma. This information relied on the self-report from each participant.

3. Participants

The final sample (n=63) was comprised of 48 females and 15 males. Ages ranged between 17 and 64 years. Over half of the participants identified themselves as being New Zealand European and a third were self-described as being Māori. Four participants identified with other ethnicities, these were British, African American, Canadian and one participant who was self described as being of other European descent. Further analysis of demographic information is presented in the results section.

4. Measures

Each participant completed an initial interview, two psychological tests (the WASI and the NART) and the NZART.

4.1. Neuropsychological measures

4.2. Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999).

The Wechsler Abbreviated Scale of Intelligence is an individually administered test of intelligence developed in 1999 in order to provide a short, reliable measure of intelligence. The WASI contains four subscales – Vocabulary, Block design, Matrix reasoning and Similarities. These subscales are similar in design to the original Wechsler Adult Intelligence Scale – third Edition (WAIS III) subscales and were chosen because they have the highest loading on g or general intelligence.

The four subtests of the WASI, which take approximately 30 minutes to administer, were selected to tap various intelligence measures such as verbal knowledge, visual information processing, spatial and non-spatial reasoning and crystallised intelligence. Together these four subscales provide a full-scale IQ score. The Vocabulary and Similarities subtests together yield the Verbal IQ score and the

Matrix Reasoning together with the Block Design subscale gives the Performance IQ score. Alternatively examiners can administer a shorter 15 minutes test comprising of only the Vocabulary and Block Design subtests. This will provide a full-scale IQ score, but does not allow for Verbal or Performance IQ scores. Details of the subtest are outlined below.

Vocabulary.

The Vocabulary subtest is comprised of 42 items. The first four items are pictures, which are shown one at a time and participants are asked to name these pictures. Items 5 - 42 are written words that are presented visually and verbally and participants are asked to verbally define their understanding of that word.

Block design.

The Block Design subtest consists of a set of 13 printed two-dimensional geometric patterns. The participant is asked to replicate the design they see within a specified time limit using coloured blocks.

Similarities.

The Similarities subtest is comprised of 22 verbal items. The Participant must explain the similarity between the two words or concepts given. The first four items are pictures and participants are asked to pick one of four items that are most similar to the other three.

Matrix reasoning.

The Matrix Reasoning subtest is a series of 35 incomplete patterns. The participant is shown a set of five possible responses and is required to choose the one they believe corresponds to complete the pattern.

4.2.1. The utility of the WASI

Historically researchers and psychologists have used short forms of the WAIS in order to obtain a quick and efficient intelligence measure. Most researchers agree a short WAIS is superior to alternative brief intelligence tests; however there are limitations to administering such a measure. Firstly, there are a number of ways to shorten the WAIS. A number of researchers have administered only some of the subscales or only a few items from each subscale based on their psychometric properties, testing time, ease of scoring, testing sequence, covering of cognitive functioning and/or clinical accuracy. Further the selection of appropriate subscales can become a very time consuming task for the researcher who has to first establish which tests are the appropriate ones to administer based on an extensive review of the literature. Finally the WAIS subtests do not have independent norms. Research has shown that individuals may perform differently when presented with two-short form tests than they do on a full WAIS.

The WASI was chosen for use in this particular research project over other tests for a number of reasons. Firstly, a short test was needed due to time constraints and also because the WASI is relatively easy to use. For example, the Wechsler Adult Intelligence Scale – Third Edition (WASI III) requires approximately 80 minutes of testing time and is comprised of 14 separate subscales. Secondly, a test developed by Wechsler was used to show some consistency with earlier research. The literature to date in regards to the NART shows a predominant use of a Wechsler test to predict premorbid IQ.

Our study also met the necessary criteria for use set out in the WASI manual. According to instructions, this test is appropriate for use in obtaining IQ scores in order to obtain estimates of current cognitive functioning and in order to obtain estimates of IQ scores for research purposes such as pre-experimental matching of cognitive ability.

4.2.2. Psychometric properties of the WASI

At the subtest level the WASI yields a reliability coefficient from .90 to .98 for vocabulary, from .84 to .96 for similarities, from .90 to .94 for block design, and from .88 to .96 for matrix reasoning. The reliability coefficients range from .92 to .98 for VIQ, from

.94 to .97 for PIQ and from .96 to .98 for FSIQ. These results are very consistent with the WAIS III.

The test-retest stability ranged from .79 to .90 for the subtests and from .87 to .92 for the IQ scales. Participants were tested twice, with a mean test-retest interval of 31 days. Inter-scorer agreement for the block design and matrix reasoning are in the high .90's. For vocabulary the score is .99 and similarities .98.

With regard to content validity the WAIS III and WASI were administered in counterbalanced order to 248 adults aged 16 – 89. The testing interval between the two administrations was from 2 to 12 weeks. The correlations were .88 for vocabulary, .76 for similarities, .83 for block design and .66 for matrix reasoning. For VIQ the correlation was .88, .84 for PIQ and .92 for FSIQ. The WASI accounts for 85% of the variance of the WAIS III and the mean IQ score of the WASI is nearly identical to the mean IQ score of the WAIS III. Further the WASI has good construct validity. Evidence of convergent and discriminative validity was based on the inter-correlations of the WASI subtests. All of the subtests correlated at least moderately with each other, ranging from the .50s to the .70s.

4.3. National Adult Reading Test (NART; Nelson and Willison, 1991)

The NART consists of a written list of 50 irregular words presented in increasing difficulty such as chord, gaoled and capon (for a full list of words see Nelson, H.E. (1982), National Adult Reading Test, Winsor, UK: NFER-Nelson). Participants are asked to read these words aloud. The words are relatively short to minimise the possible adverse effects of stimulus complexity that may occur in subjects with dementia. A NART error score is inserted into a regression equation to predict a WAIS FSIQ score. Verbal and Performance IQ scores can also be predicted using alternative equations. Full details of the NART have been presented earlier (see page 5).

4.4. New Zealand Adult Reading Test (NZART).

The NZART is based on the same methodology as the NART but based on New Zealand pronunciation and familiarity with words. The NZART consists of 60 irregular words presented in increasing difficulty (A full list of NZART words is included as Appendix B, page 71). Participants are asked to read the words aloud one at a time. The NZART error score is inserted into a regression equation to predict FSIQ, VIQ and PIQ, the latter two each having separate regression equations.

5. Procedure

Ethical approval was obtained from the Department of Psychology, University of Waikato Psychology Ethical Review Committee. Participants were recruited as described earlier. Appointments were made for participants at a time that best suited them. Each participant was assessed individually in a private interview room at the University of Waikato. Each session took approximately one and a half hours to complete.

Each session began with the initial interview where participants were given an information sheet (Appendix C), which was also explained verbally to ensure the participant fully understood the research procedure. Those participants that were still interested in taking part completed an informed consent form and a demographic form (included as Appendix D) followed by the reading of the NART, NZART and completed the WASI. The NART was read first by odd numbered participants and NZART first by even numbered participants. The NART and WASI were administered according to the instructions given in their respective manuals. In regards to the NZART, participants were asked to read aloud into a Dictaphone the NZART as best they could at a pace that they were comfortable with and were instructed to attempt words they were did not recognise. Instructions were the same for all participants. Assessments were conducted over an eight-month period, from April to November in 2005.

6. Data analysis

Participant's word pronunciation was compared to those recorded by the linguist professor. For any pronunciations that were unclear a second opinion was ought by two clinical psychology students. Data analysis was completed using SPSS (version 11.0) for windows. For each participant the number of incorrect words for both the NZART and NART were recorded, as were their FSIQ, VIQ and PIQ scores, which were calculated from the WASI. Demographic information was also recorded.

Results

1. Demographic information

Analysis of the demographic information revealed that the final sample (n=63) was comprised of 48 females and 15 males. Ages ranged between 17 and 64. The average age for males was 24.93 (standard deviation 8.10), for females the average age was slightly higher 25.08 (standard deviation 9.78).

Over half of the participants identified themselves as being New Zealand European and a third were self-described as being Māori. Four participants identified with other ethnicities these were British, African American, Canadian and one participant who was self described as being of other European descent (see Table 2).

Table 2.

Ethnicity and Gender of Participants

	Male		Female		Total	
	n	%	n	%	n	%
N.Z European	9	60.00	23	47.92	32	50.79
Māori	6	40.00	15	31.25	21	33.33
Asian	0		1	2.08	2	2.08
Pacific Islander	0		3	6.25	3	6.25
Indian	0		2	4.17	2	4.17
Other	0		4	8.33	4	8.33
Total	15	100.00	48	100.00	64	100.00

Although English as a first language was a criterion for inclusion in this study, the participants were also asked if they spoke any other languages. Almost half of the participants stated they spoke at least one other language in addition to English.

Participants were asked to record the highest education level they had attained. The majority of respondents had completed a high school qualification and only four had no formal qualification. Respondents were also asked to record their current income level (see Table 3). The number of participants in each category decreased as income increased. A third of participants had an income of less than $10,000, while only three reported an income of over $30,000.

Participants were asked about their previous mental health issues in order to clarify their suitability for undertaking the tests. Four participants had seen a psychologist in the past. Three participants had received counselling for depression and one for work related stress. However none of the participants were involved in treatment at the time of the study and three of the participants stated their mental health problems were fully resolved. Two of the participants interviewed had previously received alcohol treatment; however both had been through a rehabilitation programme some time ago and stated the matter had been settled. None of the participants were involved in treatment at the time of the study.

Participants were also asked about their previous physical health issues, again to ascertain if there were any possible confounding issues that may affect testing. It was noted that a number of participants had been hospitalised and/or injured at some point in their lives for a number of reasons not relating to head injury. None of the participants had any eyesight or reading problems that would significantly affect their ability to complete the tests. A number of participants wore corrective lenses however these were not seen as affecting their ability to undertake the tests.

1.1. Demographic differences in relation to gender

T tests revealed there were no significant differences for participants for males and females in relation to age, education, income or ethnicity. The mean age of males was 24.93 (SD = 8.102) and ranged from 17 to 45, for females the mean was slightly higher 25.08 (SD 9.780) and ranged from 17 to 61. With regard to education both male and female fell into a majority group of having attained at least a high school qualification (60% for males and 50% for females). Furthermore, the same proportion of male and female participants (50%) earned under $10,000.

The majority of both male and female participants were self-described as belonging in two majority ethnic groups, New Zealand European and Maori. For males 60% were New Zealand European and 40% were Maori. Females were 50% New Zealand European and 30% Māori.

1.2. Effect of gender on scores

There were no significant differences in the scores between males and females however, males scored higher on all WASI scales, and made fewer errors than females on both NART and the NZART (see table 3.)

Table 3

Gender differences on WASI, NART and NZART scores

	Females	Males	*t*	*df*	*p*
Full Scale IQ	100.52	104.27	1.043	61	n/s
Verbal IQ	96.83	98.47	.434	61	n/s
Performance IQ	105.04	109.67	1.415	61	n/s
NART Errors	26.54	26.33	1.272	61	n/s
NZART Errors	26.02	22.60	. 091	61	n/s

1.3. Effect of Education on scores

Analysis of the effects using Spearmans' correlation coefficient revealed a significant correlation between the qualification obtained and NART scores $r = .327$, $p <.01$ and also between the qualification obtained and NZART scores $r = .291$, $p <0.5$. Indicating that the higher the qualification obtained, the fewer the errors made on either of the reading tests.

Closer examination of these results suggested this relationship was not straight forward. Those individuals with a polytechnic qualification had the lowest mean WASI FSIQ score (89.29), VIQ score (88.57) and PIQ score (92.86). They also obtained the highest number of mean errors on the NART (31.00). However, the private training group obtained the highest mean number of errors on the NZART (32.33).

Those students with an honours degree obtained highest mean FSIQ (116.40) and VIQ (117.80). This group also scored the lowest mean number of NZART errors (11.20). The lowest mean number of errors for the NART was obtained by the masters group (12.00), and the highest mean PIQ score (113.20) was found in the diploma group.

The highest overall number of errors on the NART (n=43) and NZART (n=46) was scored by those participants who reported their highest qualification as a high school qualification. The lowest overall number of errors for both NART (n=8) and NZART (n=7) was scored by the honours degree group.

1.4. Effect of ethnicity on scores

Of the two majority groups there were more European (n=32) than Māori (n=21). Examination of the differences between the scores obtained on the WASI and the two reading tests revealed that on average New Zealand Europeans scored higher on all WASI scales than Māori and made fewer errors overall on both NART and NZART. However, both Māori and New Zealand Europeans made more errors on average on the NART than on the NZART (see table 4).

Further analysis indicated there was a significant difference between the two groups in relation to FSIQ and VIQ scores, European New Zealanders scored significantly higher in both WASI FSIQ and VIQ, however no significant differences were found for PIQ, NART or NZART scores.

Table 4

Differences in scores for Māori and Europeans

	Māori	NZ European	*t*	df	*p*
FSIQ	97.24	104.97	2.590	51	*p<.05*
VIQ	103.81	108.91	2.245	51	*p<.05*
PIQ	93.05	100.38	1.883	51	n/s
NART	28.76	25.88	1.320	51	*n/s*
NZART	27.90	23.56	1.696	51	*n/s*

1.5. Effect of income on scores

Participants were split into two groups those earning over $10,000 (n=26) and those earning under $10,000 (n=32). No significant differences were found between these two groups on any of the test scores. Furthermore no significant correlation was found between any of the obtained scores and income level.

The highest mean FSIQ score obtained was 113.33 from the group earning over $30,000 this group also scored the highest mean VIQ score (107.67), the highest PIQ score (116.00) and the lowest mean number of errors for both NART (11.00) and NZART (10.67).

The lowest mean FSIQ score was obtained by the $10,000 to$ 20,000 income group (98.29); this group also scored the lowest PIQ (103.06) and VIQ scores. (94.00) and the highest number of mean errors on the both the NART (29.50) and NZART (29.00).

2. Correlations between WASI, FSIQ, VIQ and PIQ scores

A Pearsons correlation was conducted to determine the association between the

scores on the different tests. The correlation coefficient revealed that all scores were at least moderately correlated. For example, PIQ correlated moderately with both the NZART (r = .439, p <.001) and NART (r = .411, p <.001). FSIQ scores also showed a moderate correlation with the NZART (r = .679, p <.001) and NART (r = .650, p < .001).

Other scores were more highly correlated. The NART and NZART had a correlation of r = .917, (p <.001) and VIQ scores were also highly correlated with both the NZART (r =.741, p <.001) and NART (r =.703, p <.001).

3. Regression equation for a New Zealand population

A linear regression was carried out with the NART error score and WAIS FSIQ, VIQ and PIQ. The regression equations produced are represented below:

WAIS FSIQ = 128.775 - (1.033 x NART error)

WASI VIQ score = 128.019 - (1.162 x NART error)

WASI PIQ score = 121.987 - (.598 x NART error)

The regression equation predicted 42% of variance in WASI FSIQ, 49% in WASI VIQ, and 17% in WASI PIQ.

In order to create comparative data in the accuracy of IQ the WASI FSIQ, VIQ and PIQ were separated in ranges spanning 10 points each. The average WASI FSIQ, VIQ and PIQ were then calculated for both obtained and predicted IQ scores and these scores were compared graphically. The obtained and predicted WASI FSIQ, VIQ and PIQ scores are shown in Figures 1, 2 and 3 (these can be found on page 44).

The regression equation for WASI FSIQ overestimates FSIQ by 22 points on average at the lower end of IQ range and underestimate IQ at the higher end by 11.5 points. The estimated IQ scores are more accurate in the middle ranges with an average difference of 0.24 points overall. The most accurate being at the 100 – 109 mark where the equation underestimates FSIQ by an average of 1.87 points. Overall the New Zealand equation overestimates FSIQ by 1.59 points on average.

The predicted WASI VIQ and PIQ scores follow a similar trend. They underestimate in the lower ranges, overestimate in the higher ranges and are more accurate towards the middle ranges. The average difference in the middle ranges for

VIQ is 2.04 points. The lowest VIQ range is overestimated on average by 10.25 points and underestimated in the highest VIQ range by 9.75 points. The predicted VIQ equation is most accurate at the 90 – 99 range where VIQ is overestimated on average by 0.7 point. Overall VIQ is on average, underestimated by 1.28 points.

Similarly WASI PIQ is overestimated by 17.66 points in the lowest range and underestimated at the higher end by 11.57 points. Again the prediction is closer to the obtained score in the middle ranges, the average difference is 0.17 points. Overall the regression equation underestimates PIQ by 1.11 points on average.

4. Original NART

In order to establish the accuracy of the New Zealand normed regression equation a comparison with the British regression equation was necessary. The British regression equation was used to generate graphs using the same method as for the New Zealand normed equation. A graphical analysis of the obtained WASI FSIQ, VIQ and PIQ scores and predicted scores based on the original NART are included as Figure 4, 5 and 6 (refer to page 45).

A similar trend was revealed between the New Zealand NART and the original NART. The Original NART also overestimated IQ in the lower ranges and underestimated IQ in the higher ranges. For WASI FSIQ the equation overestimated IQ on average by 17.75 points and underestimated the IQ at the highest range by 12 points. In the middle ranges the IQ score was again more accurate with the same average difference as the New Zealand normed equation of 4.28 points, this is much higher than in the New Zealand normed equation. Furthermore in contrast to the New Zealand normed scores the original NART equation continued to underestimate FSIQ at all other ranges with the exception of the 90 – 99 range, which is an exact prediction. Overall the British equation underestimated FSIQ on average by 1.89 points.

The original equation for WASI VIQ follows a very similar trend to that obtained using a New Zealand population. At the lowest range VIQ is overestimated by 11.75 points on average and underestimated by 10.25 points in the highest range. Again the most accurate range is the 90 – 99 range were the equation overestimates by 0.43. The average difference in the middle ranges of 2.28 is also very similar to the average difference obtained for the New Zealand normed equation. Overall the British regression equation is the same as that normed on a New Zealand population and underestimates VIQ on average by 1.28 points.

For WASI PIQ the equation overestimates IQ at the lowest range on average by 7 points and underestimates in the highest range by 14.1 points. These results are quite different to those found in the New Zealand normed equation. The British NART shows an average difference of 8.2 in the middle ranges. The overall average difference for PIQ is 6.37 points. The results for the British normed NART underestimate IQ at a much greater level than that of the New Zealand normed IQ. For PIQ the IQ level is underestimated at all ranges except the first.

5. New Zealand Adult Reading Test

A linear regression was carried out with the NZART error score and WAIS FSIQ, VIQ and PIQ. The regression equations produced are represented below:

WASI FSIQ score = 124.18 - (.903 x NART error)

WASI VIQ score = 123.069 - (1.025 x NART error)

WASI PIQ score = 119.616 - (.535 x NART error)

This regression equation predicted 46% of variance in WASI FSIQ, 55% in WASI VIQ, and 19% in WASI PIQ. A graphical analysis of obtained and predicted NZART scores is presented in figures 7, 8 and 9.

Results for the WASI FSIQ scores were very similar to both previous equations. The predicted FSIQ score was overestimated by 21.25 points at the lowest range and underestimated by 12.25 points at the highest range. Once again the equation was more accurate in the middle with an average difference of 0.37, this is slightly lower than the British normed NART, but higher than the New Zealand normed NART. Overall the NZART error equation has a better prediction than both other regression equations (1.26).

The predicted WASI VIQ score was overestimated in the lower ranges by 9.75 points and under estimated by 9.25 points, again these results are reasonably consistent with those found in the two earlier regression equation. The predicted VIQ equation was again most accurate at the 90 – 99 range where the prediction overestimates by 0.2. In the middle ranges the average difference was slightly higher than both the previous equations (1.43) and overall the NZART error equation underestimates VIQ by 0.87 points.

For WASI PIQ prediction the results showed a similar trend. At the lower range the equation overestimated by 16.98 points and at the higher range the prediction was underestimated by 12.28 points. The equation is almost accurate in the middle ranges being only .04 overestimated. Furthermore, the NZART regression equation for PIQ is on average more accurate overall than both the British and the New Zealand normed equation where PIQ is overestimated by (.96)

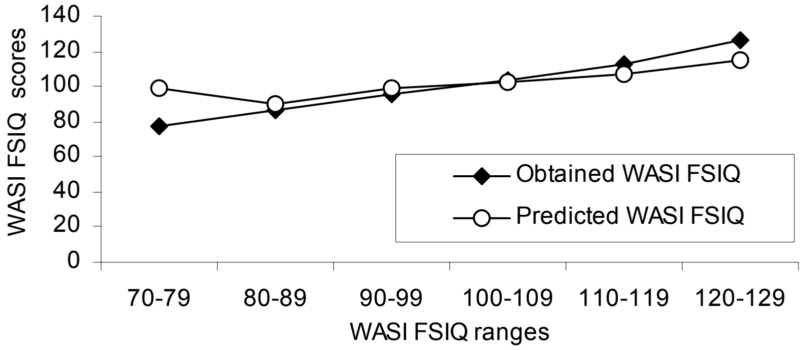

Figure 1. Predicted and obtained WASI FSIQ scores based on the regression equation normed on a New Zealand population.

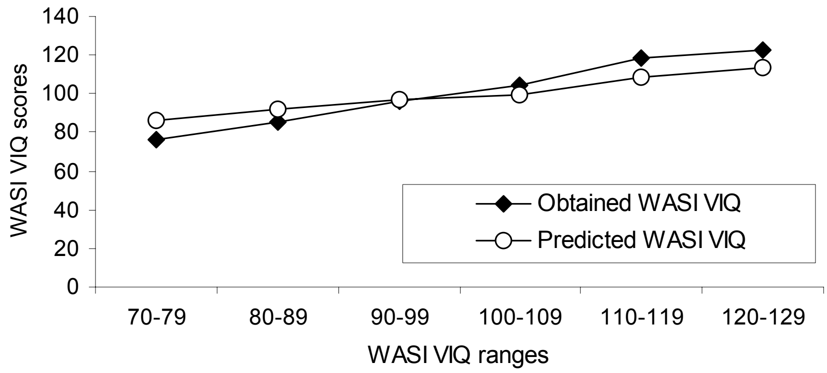

Figure 2. Predicted and obtained WASI VIQ using a regression equation based on a New Zealand population.

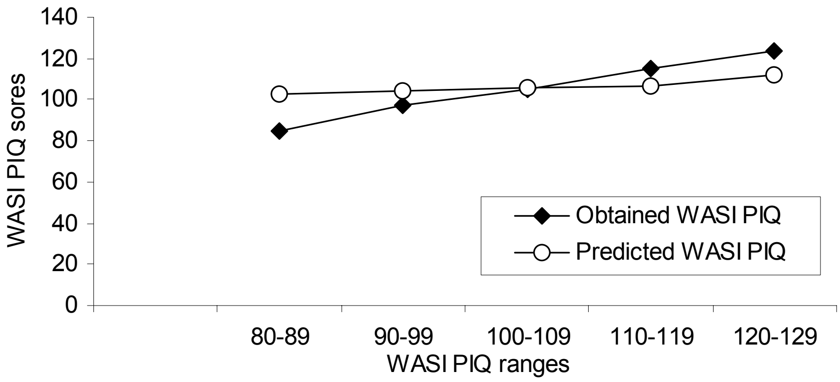

Figure 3. Predicted and obtained WASI PIQ scores based on a regression equation normed on a New Zealand population.

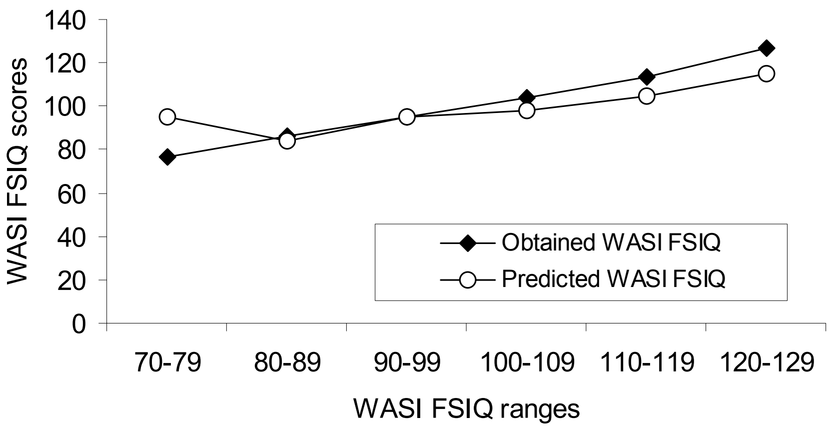

Figure 4. Predicted and obtained WASI FSIQ scores based on the regression equation from the original NART.

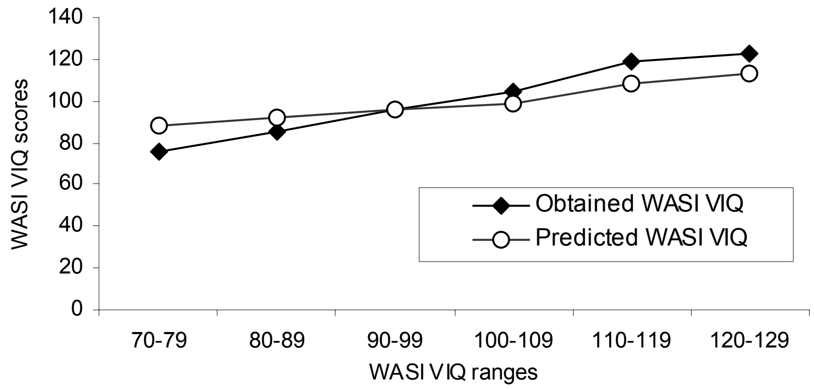

Figure 5. Predicted and obtained WASI VIQ based on the original NART regression equation.

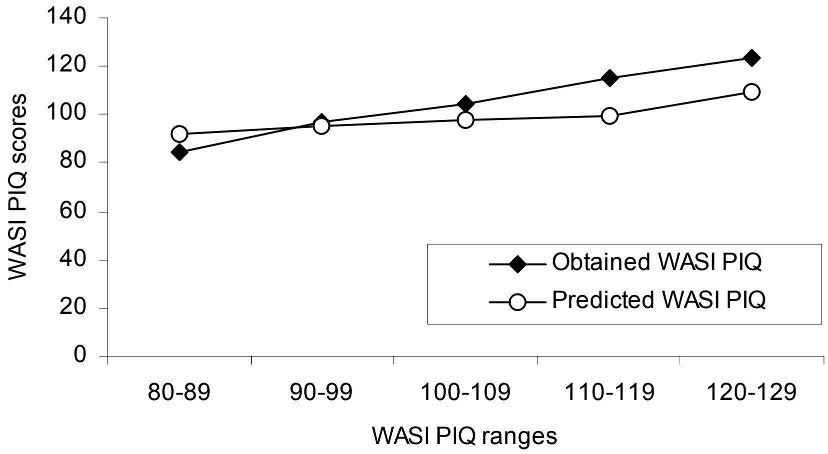

Figure 6. Predicted and obtained WASI PIQ based on the original NART regression equation.

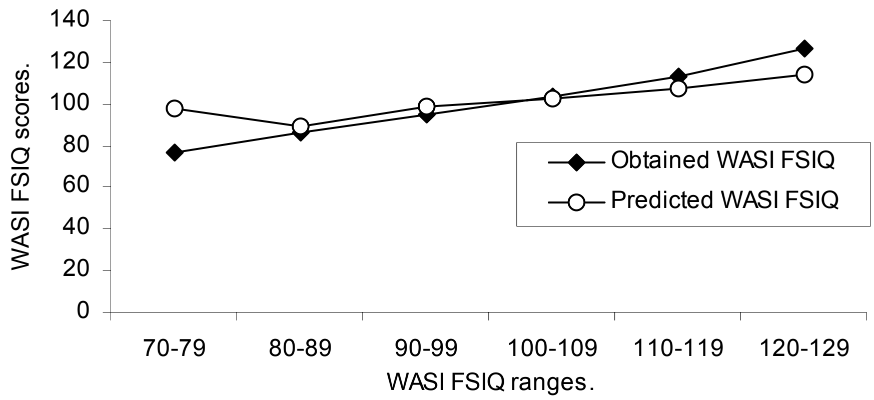

Figure 7. Obtained WASI FSIQ scores and predicted WASI FSIQ scores based on a regression equation from NZART errors.

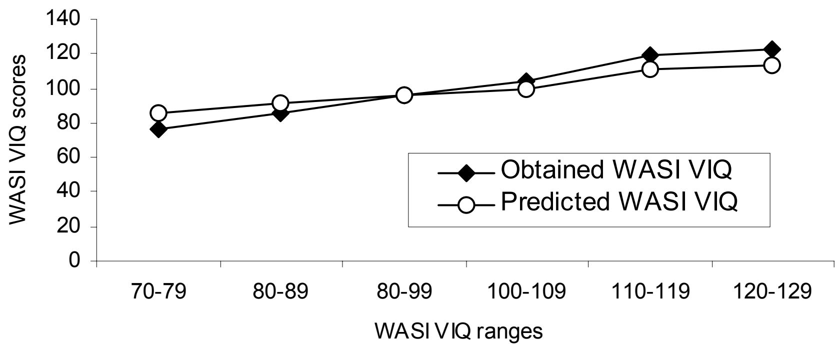

Figure 8. Obtained VIQ scores and predicted VIQ scores based on a regression equation from NZART errors.

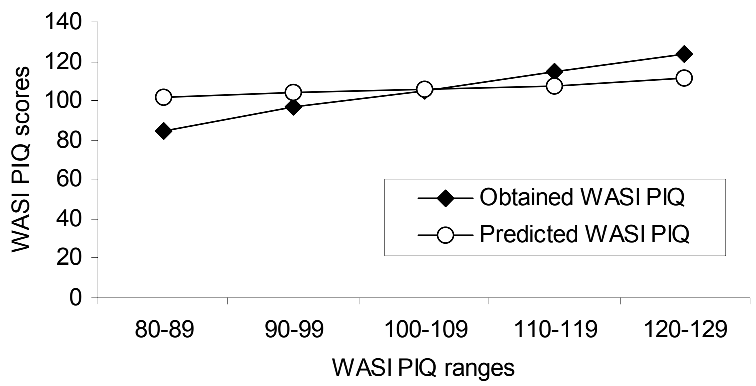

Figure 9. Obtained WASI PIQ scores and predicted WASI PIQ scores based on a regression equation from NZART errors.

Discussion

The current experiment had three aims. Firstly, to ascertain whether demographic variables have a significant effect on the NART, WASI or NZART scores in a New Zealand population. Secondly, to develop New Zealand norms for the existing NART and finally to develop a New Zealand Adult Reading Test that better predicts premorbid IQ for a New Zealand population.

1. Demographic variables

1.2. Age, income level and gender

There was no significant difference between gender, age and income and the WASI, NART and NZART scores obtained. These variables were found to be reasonably robust measures and relatively resistant to testing. This is consistent with previous research with both the original NART and the revised versions of the NART. Most researchers agree that age, gender and social class are not relevant factors in determining IQ (Nelson, 1978; Crawford, Stewart & Garthwaite et al., 1988; Blair & Spreen, 1989, Nelson & Willison, 1991; Wiens et al., 1993; Schlosser & Ivison, 1989; Fisher, 1996; Uttl, 2002). However, a cautionary note needs to mentioned here; namely that previous researchers used the category social class rather than income level, and the two may not be interchangeable.

A more detailed examination of the WASI, NART and NZART scores revealed that males scored higher than females on all WASI scales, on average males scored 3.75 points higher than females on the WASI FSIQ, 1.64 points higher on the WASI VIQ and 4.63 points higher on the WASI PIQ. Furthermore males made fewer errors than females on both the NART and NZART. Males made 3.42 less errors on average with the NZART and 0.21 fewer errors on the NART, although interesting, these differences were not significant and is inconsistent with other research findings. Wiens et al (1993) found that females made fewer errors on the NAART than males; a finding, which was also replicated by Uttl (2002). In a New Zealand study Fisher (1996) also found males had a higher error score than females on the NART. It is unclear why the results differ from previous research. However, it could be postulated that the sample used here is relatively different from those used in previous research. Both the Wiens et al (1993) and Uttl (2002) studies were completed in North America and used the NAART. Gender and age did not differ greatly from the present study; however, ethnicity or education may be contributing factors.

Fisher's (1996) research included fewer Māori than the present study and no analyses were made to ascertain whether there were differences between the genders in regard to ethnicity and scores. Furthermore, although Fisher's (1996) study was conducted in the same general demographic area as the present study, it was not conducted with university students, and this may have had an effect on the obtained scores.

1.3. Ethnicity and education

Examination of the effects of ethnicity on results revealed that there was a significant difference between the two ethnic groups in relation to test scores. However, this was found only in regard to WASI FSIQ and VIQ scores. The results suggest there may be something particular to the WASI FSIQ and VIQ that disadvantages Māori. Both of these tests are highly loaded in verbal ability, while PIQ samples spatial reasoning and practical ability. Previous research has shown verbal item content does affect testing (Barker-Collo, 2001), and that language is influenced by culture (Harvey & Siegert, 1999). It may be that Europeans are more familiar with the verbal items in the WASI tests than Māori.

However to suggest the difference is solely due to verbal ability would not explain why ethnicity did not affect the NART and NZART scores. It can only be hypothesized that Māori were more familiar with NART and NZART words or that they found these tests easier than the WASI VIQ. This could possibly be due to the demands of the VIQ subscale. The NART requires the reading of single words, while the subscale of the WASI requires abstract thinking and verbal definitions. Further investigation into this suggestion would be valuable

Researchers have shown Māori perform at a significantly lower level than non-Māori in psychometric testing involving verbal ability such as the Boston Naming Test and the Rey Auditory Verbal Learning Test (Barnfield & Leathem, 1998; Barker-Collo, 2001). This finding held true for our study, although the difference in scores was not significant, New Zealand Europeans made fewer mean errors overall on both the NART and NZART and scored more highly on all WASI scales. Although some of the words included in the NZART were Māori words the majority of words were English. This finding further supports the idea that word content and culture affects testing ability.

Interestingly, both Māori and New Zealand Europeans made fewer errors on the NZART than on the NART. This could suggest the NZART is easier than the NART or again provide further support that the participants were more familiar with the NZART words. Research does suggest that New Zealand participants have difficulty with word items they are not familiar with such as sweater and beaver (Rolleston, 2001; Barker-Collo, 2001).

As expected there was a correlation between qualification obtained and NART and NZART scores. This is consistent with the research conducted by Crawford, Stewart, Garthwaite et al (1988) and Fisher (1996). Researchers have suggested a higher level of education implies a higher level of verbal ability and thus a higher score (Crawford, 1989; Barker-Collo, 2001). This was certainly true in the current experiment. Those individuals who had attained a university degree had a higher mean WASI score across all three WASI scales, although this was not significant. There was however, a statistically significant difference in the number of mean errors in both the NART and NZART error scores

Results were less clear with the lower scoring groups. The participants who obtained a polytechnic qualification had the lowest WASI scores across all three subtests, and the highest number of errors on the NART. However, it was the group who had obtained a private training qualification who recorded the lowest error score on the NZART.

Overall, as expected these results suggest the higher the qualification, the higher the score on the WASI scales and the lower the number of errors on the NZART and NART. With a lower level of education results seem to vary. However it is difficult to quantify the differences between the categories used in this study; for example, between the private training group and the polytechnic group. In retrospect, years of education may have been a more flexible category.

In summary, the current experiment reiterated the findings of previous research; that age, gender and income level did not significantly affect NART or IQ scores, while ethnicity and culture did impact on the scores. This was especially where verbal content is concerned. The present study also adds weight to the concerns expressed by researchers in relation to the use of tests for a cultural group other than the one it was developed for. Furthermore the current study demonstrated that education is correlated with and affects IQ scores.

2. Development of New Zealand norms for the NART

2.1. Full Scale, Verbal and Performance IQ

The second aim of this thesis was that by establishing norms for a New Zealand population the original NART would be a better predictor of premorbid IQ for a New Zealand sample. Examination of the results revealed this hypothesis was partially supported.

Overall there was little difference between the accuracy of the British NART and the New Zealand normed NART, in regard to WASI FSIQ and VIQ. The New Zealand normed NART is slightly superior in the higher ranges, while the British normed NART is slightly superior in the lower ranges. Both equations consistently overestimated IQ at the lower ranges of IQ and underestimated IQ at the higher ranges. This is consistent with earlier research by Fisher (1996) where results showed the NART overestimated WAIS-R FSIQ at the lower end of the IQ range by 17 points and underestimated the WAIS-R FSIQ at the higher end by 18 points. By way of comparison, Fishers' (1996) study underestimated IQ by a much larger point average than the current study. It may be that discrepancies at the extreme ends of the range are exacerbated by the small group size, as there were only a few scores represented at the lowest range, however this would not account for the discrepancy at the higher ranges.

Other researchers have also noted the inability of the NART to accurately estimate the whole range of premorbid IQ; with reports of consistent overestimation in the lower ranges and underestimation in the higher ranges (Fisher, 1996; Wiens, et al., 1993). However, the intention of the NART is not to precisely predict individual IQ scores but to provide a relatively accurate estimate of IQ applicable for a large population. Given that the overestimation in prediction recorded by the New Zealand normed NART was only 1.59 points for WASI FSIQ, it can be concluded that it achieves this goal.

In regard to WASI PIQ, the British NART results were in very different to the New Zealand normed results. The British NART underestimated PIQ at every range except the very lowest range, and had a much higher mean average difference at both the mid range and overall. This indicates that the NART normed on a British population is not a very good predictor of performance IQ in a New Zealand sample. This again reflects the cultural aspect of testing and adds weight to the bias found in psychometric testing. It is proposed here that the reason for such a large underestimation in WASI PIQ reflects the difference in emphasis placed by the two relevant cultures on practical ability. The British normed PIQ predicts a much lower performance score than that obtained by the New Zealand participants.

2.2. The regression equation

The regression equation for FSIQ calculated from New Zealand norms predicted 42% of variance in FSIQ. This is much lower than previous research such as Nelson & O'Connell (1978), where FSIQ accounted for 55% of variance and Crawford, Parker, Stewart et al (1989) where 66% of variance was accounted for.

The New Zealand normed VIQ and PIQ results also accounted for a much lower variance when compared with previous overseas studies. The regression equation for VIQ in the New Zealand normed sample accounted for 49% of variance; this is compared to Nelson & O'Connell (1991) (VIQ = 60%), and Crawford, Parker, Stewart et al (1989), (72% VIQ). The New Zealand normed regression equation accounted for 17% of variance in PIQ. Again, this is much lower than those recorded by Nelson & O'Connell (1978), and Crawford, Parker, Stewart et al (1989) where variance accounted for was recorded as being in the low 30's.

The above comparisons indicate that these reading tests do not predict IQ in New Zealand as well as they have in other countries. That is, less variance in IQ is explained by the NART in a New Zealand population, than that which has been explained in other countries. The reasons for such discrepancies are not entirely clear. There appears to be little overall difference between the present study and other research in regard to demographic variables, with the exception of culture, which may be a contributing factor. Samples sizes of previous research were recorded as 120 and 151, which may also be a contributing factor. Crawford, Parker, Stewart et al (1989) suggests that by using a larger sample size, variance accounted for may be enhanced due to the wider variation in IQ scores.

3. The development of a New Zealand Adult Reading Test

The third and final hypothesis of this thesis was that the NZART would be more accurate than both the New Zealand normed NART and the original NART in predicting IQ. The results indicate that this hypothesis was indeed supported. The regression equation for the NZART was a more accurate predictor across all three subtests of the WASI than either of the NART equations.

At the lower end of the IQ range, the NZART overestimated WASI FSIQ by a slightly higher margin than the original NART, but was comparable to the New Zealand normed equation. The British NART appeared to be slightly superior at the lowest ranges but overall the NZART yielded more accurate WASI FSIQ estimate.

As expected the NZART was a better measure of WASI VIQ than either of the equations developed using the NART. The NZART was also superior at all ranges and gave a more accurate overall WASI VIQ. The same general trend was found for the WASI PIQ. Although the British equation was slightly superior at the lowest ranges, overall the NZART was more accurate.

3.1. The regression equation

Given the results of comparison between previous regression equations and the New Zealand ones, it was not surprising to find that the NZART regression equation explained less variance than those equations developed overseas. The NZART regression equation predicted 46% of variance in WASI FSIQ, 55% in WASI VIQ, and 19% in WASI PIQ, much lower than most other studies.

The variance accounted for in both the AMNART (Swartz & Saffran, 1987) and NAART (Blair & Spreen, 1989) are very similar to each other. They are also very comparable to those found by Nelson & O'Connell (1991) and Crawford, Parker, Stewart et al (1989). However, the sample size for the AMNART and NAART studies were much smaller than those used by Nelson & O'Connell (1991) or Crawford, Parker, Stewart et al (1989), being 109 and 66 respectively; these are comparable to the current study and suggests sample size was not a relevant factor in explaining the difference.

It is therefore postulated that perhaps New Zealand places a higher emphasis on practical skills and abstract reasoning rather than reading ability. This would explain why the IQ scores are less well explained by ability to read, and further explain why Māori score lower on these tests than Europeans.

Further testing in the future would ascertain if this is the case. However a confounding variable that must be considered is the IQ test itself. Previous researchers have used the full or shortened version of the WAIS; in contrast the current research used the WASI. It is always possible that the prediction may be enhanced with the use of the WAIS-III; again, future testing in this regard would be valuable.

In conclusion, the British NART underestimated IQ in all three subscales of the WASI. In contrast the NZART and New Zealand normed equation overestimated PIQ and FSIQ and underestimated VIQ. However with the exception of the British normed equation's estimation of PIQ, these discrepancies were relatively minor. Thus it is concluded that this study showed the NZART as a superior measure of premorbid IQ in a non-clinical sample. This re-emphases the need to both develop psychometric test relevant to the country and cultures they are intended to be used with. Furthermore this study also showed the New Zealand normed regression equation as superior to the British equation, again re-emphasising the need for caution when using overseas test.

4. Methodological issues and implications

As with any research, a number of concerns can be raised over validity, reliability and the application of the results to real life situations. The present study was conducted with university students, most of whom were first year students with a similar age range and this group is unlikely to be representative of the general population.

Of particular concern is the small sample size and limited representation shown of some groups. Although little difference was found between our groups in regard to gender, education, income, age and ethnicity, some of these subsets had very small numbers. Future research involving larger representation would show if results remain constant when the dynamics of these groups changes.

Another possible concern could be raised over the differences in categorisation between this study and those used in previous research. In the present experiment income level and educational attainment were used. Wiens, et al (1993) and Fisher, (1996) use occupational categories, while other researchers (Nelson & Willison, 1991; Crawford, Stewart, Garthwaite et al., 1988) have used class and based their class categorisation on their particular countries. Education, on the other hand has been categorised by previous researchers as years of schooling. It is always possible the categories are very different and perhaps the class system is not interchangeable internationally. It is also possible that an individual's years of education yield a more accurate result than educational attainment. In the present study it is difficult to ascertain where each category lies in regard to quantifying education. Therefore using a more consistent measure may yield different results, and enable direct comparisons to be made with previous research.

Although it has been well documented that the NART is a better predictor of IQ in the middle ranges (Nelson & Willison, 1991; Wiens, et al., 1993; Fisher, 1996). This finding was certainly validated in the present experiment. It is worth mentioning the possibility of ceiling effects. An individual who scores 0 errors on the NZART has an equivalent IQ of 124, in comparison an individual with the same score will have an IQ score of 129 in the New Zealand normed NART and an IQ of 131 with the British normed NART (Test data for the NZART and New Zealand normed NART can be found in appendices E and F on pages 75 and 77). This indicates a ceiling effect that one would have to consider when testing if results were to show an individual had an extremely high IQ score.

The same cautionary note is needed when testing reveals an extremely low IQ score, however this would only be relevant for WASI PIQ; a maximum error score on the NZART yields a FSIQ of 70 and a VIQ of 62, suggesting the individual may have an intellectual disability, and that this test is not a suitable measure. However an individual with an absolute 100% error score on the NZART would obtain a PIQ score of 88. By way of comparison, the same score would yield 92 on the New Zealand normed NART and 73 with the original NART.

However when one considers the majority of IQ scores will invariably fall between 70 and 130, and that the purpose of the NZART is to primarily establish Full Scale IQ, there should be little concern with the NZART having a floor effect or ceiling effect.

5. Implications for future research

The importance of developing psychometric tests for use in the population they were normed on cannot be overstated. Researchers have consistently cautioned against using tests developed overseas. The development of a NZART is a step in this direction, however the NZART is far from perfect. There is a need to further develop the NZART. This could be achieved by testing different words to modify the original NZART. Further studies could include different sample groups and a larger sample population aimed at achieving a wider representation of minority groups and wider variation in IQ scores. In essence these factors may enhance the accuracy of the NZART.

Furthermore the NZART is yet to be tested on a clinical sample to validate its applicability and reliability. In order to test the NZARTs' reliability and validity there is a need to conduct trials with a wide range of clinical populations. Inter-rater reliability checks, test -retest, and split half reliability tests would also enhance test validity, as would predictive validity tests based on prior records of cognitive ability such as academic or military records.

Finally the NZART is quite long, and given the population it is intended for, it may cause individuals to become tired, thus impairing performance. A shortened version of NZART could be a developed as a viable option.

6. Conclusion

The results presented in this study show that demographic variables such as age, gender and income have no effect on the prediction of IQ, while ethnicity and education are significantly correlated. Furthermore, this experiment showed the NZART as a better predictor of premorbid IQ than the NART normed on a New Zealand population and the original British NART. In essence this finding supports the development of psychometric tests by the country and culture it is intended to be used with. Future research is needed to enhance the accuracy of the NZART and establish its' reliability and validity among both clinical and unimpaired sample groups.

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing.* Prentice-Hall International.

Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17,* 143-150.

Ardila, A., Rosselli, M., & Puente, A.E. (1994). *Neuropsychological evaluation of the Spanish speaker.* New York, Plenum Press.

Barker-Collo, S.L. (2001). The 60-item Boston naming test: Cultural bias and possible adaptations for New Zealand. *Aphasiology, 15,* 85-92.

Barker-Collo, S.L., Clarkson, A., Cribb, A., & Grogan, M. (2002). The impact of American content on California Verbal Learning Test performance: A New Zealand illustration. *The Clinical Neuropsychologist, 16,* 290-299.

Barker-Collo, S.L. (2003). Culture and validity of the Symptom Checklist-90-Revised and Profile of Mood States in a New Zealand sample. *Cultural Diversity and Ethnic Minority Psychology, 9*, 185-196.

Barnfield, T.V., & Leathem, J.M. (1998). Neurological outcomes of traumatic brain injury and substance abuse in a New Zealand prison population. *Brain Injury, 12,* 951-962.

Barona, A., Reynolds, C. R., & Chastain, R. L. (1984). An improved estimate of premorbid IQ for blacks and whites on the WAIS-R. *International Journal of Clinical Psychology, 8,* 169-173.

Berry, D.T.R., Carpenter, G.S., Campbell, D.A., Schmitt, F A., Helton, K., & Lipke-Molby, T. (1994). The New Adult Reading Test – Revised: Accuracy in estimating WAIS-R IQ scores obtained 3.5 years earlier from normal older persons. *Archives of Clinical Neuropsychology, 9,* 239-250.

Blair, J.R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist, 3,* 129-136.

Boekamp, J.R., Strauss, M E., & Adams, N. (1995). Estimating premorbid intelligence in African Americans and white elderly veterans using the American version of the National Adult Reading Test. *Journal of Clinical and Experimental Psychology, 17,* 645-653.

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist, 44,* 567–578.

Cockburn, J., Keene, J., Hope, T., & Smith, P. (2000). Progressive decline in NART score with increasing dementia severity. *Journal of Clinical and Experimental Neuropsychology, 22,* 508-517.

Crawford, J.R. (1989). Estimation of premorbid intelligence: A review of recent developments. In J. R Crawford & D. M. Parker (Eds.), *Developments in clinical and experimental neuropsychology (p55-74).* New York: Plenum.

Crawford, J.R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J.R. Crawford, D.M. Parker & W.W. McKinlay (Eds.), *A handbook of neuropsychological assessment.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Crawford, J.R., & Allan, K.M. (1997). Estimating premorbid IQ with demographic variables: Regression equations derived from a U.K. sample. *Clinical Neuropsychologist, 27,* 181-182.

Crawford, J.R., Allan, K.M., Cochrane, R.H.B., & Parker, D.M. (1990). Assessing the validity of NART estimated premorbid IQ's in the individual case. *British Journal of Clinical Psychology, 29,* 435-436.

Crawford, J.R., Besson, J.A.O., Bremner, M., Ebmeier, K.P., Cochrane, R.H.B., & Kirkwood, K. (1992). Estimation of premorbid intelligence in schizophrenia. *British Journal of Psychiatry, 161,* 69-74.

Crawford, J.R., Besson, J.A.O., Parker, D.M., Sutherland, K.M., & Keen, P.L. (1987). Estimation of premorbid intellectual status in depression. *British Journal of Clinical Psychology, 26,* 312-314.

Crawford, J.R., Deary, I.J., Starr, J., & Whalley, L.J. (2001). The NART as an index of prior intellectual functioning: A retrospective validity study covering a 66-year interval. *Psychological Medicine, 31,* 451-458.

Crawford, J.R., Parker, D.M., & Besson, J.A.O. (1988). Estimation of premorbid intelligence in organic conditions. *British Journal of Psychiatry, 153,* 178-181.

Crawford, J.R., Parker, D.M., Stewart, L.E., Besson, J.A.O., & De Lacy, G. (1989). Prediction of the WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology, 28,* 267-273.

Crawford, J.R., Stewart, L.E., Cochrane, R.H.B., Foulds, J.A., Besson, J.A.O., & Parker, D.M. (1989). Estimating premorbid IQ from demographic variables: Regression equations derived from a UK sample. *British Journal of Clinical Psychology, 28,* 275-278.

Crawford, J.R., Stewart, L.E., Cochrane, R.H.B., Parker, D.M., & Besson, J.A.O. (1989). Construct validity of the National Adult reading Test: A factor analytic study. *Personality and Individual Differences, 10*, 585-587.

Crawford, J.R., Stewart, L.E., Garthwaite, P.H., Parker, D.M., & Besson, J.A.O. (1988). The relationship between demographic variables and NART performance in normal subjects. *British Journal of Clinical Psychology, 27,* 181-182.

Crawford, J.R., Venneri, A., & O'Carroll, R.E. (1998). *Comprehensive Clinical Psychology: Geropsychology Vol 7. Clinical neuropsychological assessment of the elderly.* Retrieved June 10, 2005, from http://www.sciencedirect.com.ezproxy.waikato.ac.nz:2048/science/reference works/0080427073

Cubelli, R. (2005). The history of neuropsychology according to Norman Geschwind: Continuity and discontinuity in the development of science. *Cortex, 41*, 271– 274.

Deverson, T., & Kennedy. (Eds.). (2005). *The New Zealand Oxford dictionary.* Auckland: Oxford University Press.

Doyle. K. O. (1974). Theory and practice of ability testing in ancient Greece. *Journal of the History of the Behavioral Sciences. 10,* 202-212.

Ebmeier, K.P., Booker, K., Cull, A., Gregor, A., Goodwin, G.M., & O'Carroll, R.E. (1993). The validity of the National Adult Reading Test in estimating premorbid intellectual ability in long-term survivors of hemispheric glioma and whole brain irridation – A pilot study. *Psycho-Oncology, 2,* 133-137.

Eppinger, M.G., Craig, P.L., Adams, R.L., & Parsons, O.A. (1987). The WAIS-R index for estimating premorbid intelligence: Cross-validation and clinical utility. *Journal of Consulting and Clinical Psychology, 55,* 86-90.

Eubanks, J.D. (1997). Division 40 report: Clinical neuropsychology summary information prepared by division 40, clinical neuropsychology, American Psychological Association. *The Clinical Neuropsychologist, 11,* 77-80

Fanzen, M.D., Burgess, E.J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology, 12,* 711-738.

Fisher, J.M. (1996). *Estimating premorbid intelligence with the National Adult Reading Test in New Zealand*. Unpublished masters thesis. University of Waikato, New Zealand.

Freeman, J., Godfrey, H.P.D., Harris, J.K.J., & Partridge, F.M. (2001). Utility of a Demographic equation in detecting impaired NART performance after TBI. *British Journal of Clinical Psychology, 40,* 221-224.

Fromm, D., Holland, A.L., Nebes, R.D., & Oakley, M.A. (1991). A longitudinal study of word-reading ability in Alzheimer's disease: Evidence from the National Adult Reading Test. *Cortex, 27,* 367-376.

Griffin, S.L., Mindt, M.R., Rankin, E.J., Ritchie, A.J., & Scott, J.G. (2002). Estimating premorbid intelligence comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology, 17,* 497-507.

Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology, 13*, 933-949.

Hart, S., Smith, C.M., & Swash, M. (1986). Assessing intellectual deterioration. *British Journal of Clinical Psychology, 25,* 119-124.

Harvey, J.A., & Siegert, R.J. (1999). Normative data for New Zealand elders on the Controlled Oral Word Association Test, Graded Naming Test, and the Recognition Memory Test. *New Zealand Journal of Psychology, 28,* 124-140.

Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47*, 1083-1101.

Heriot, S., & Beale, I. (1996). New Zealand norms for the Visual Sequential Memory Subtest of the Illinois Test of Psycholinguistic Abilities-Revised. *New Zealand Journal of Psychology*, *25,* 2-6.

Issella, V., Villa, M.L., Forapani, E., Piamarta, F., Russo, A., & Appollonio, I.M. (2005). Ineffectiveness of an Italian NART-equivalent for the estimation of verbal learning ability in normal adults. *Journal of Clinical and Experimental Neuropsychology, 27,* 618-623.

Johnstone, B., Callahan, C.D., Kapila, C.J., & Bouman, D.E. (1996). The comparability of the WRAT-R Reading Test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Archives of Clinical Neuropsychology, 11,* 513-519.

Klesges, R.C., Fisher, L., Vasey, M., & Pheley, A. (1985). Cross-validation of an adult premobid functioning index. *Clinical Neuropsychology, 3,* 13-15.

Kurtzke, J.F. (1984). *Neuroepidemiology. Annals of neurology*, 16, 265-277.

Lezak, M. (1995). *Neuropsychological assessment.* (3rd Ed.). Oxford UK: Oxford University Press.

Liangshi, Y. (1999). Ancient Chinese mental tests and the features and values. *Psychological Science (China). 22,* 132–135.

MacKinnon, A., & Mulligan, R. The estimation of premorbid intelligence levels in French speakers/Estimaton de l'intelligence prémorbide chez les francophores. *Encephale, 31,* 31-43.

McFie, J. (1975). *Assessment of organic intellectual impairment*. New York: Academic Press. .

Mortensen, E.L., Glade, A., & Reinisch, J.M. (1991). Best performance method in clinical neuropsychology. *Journal of Clinical and Experimental, 13,* 361–371.

Moss, A. R., & Dowd T. (1991). Does the NART hold after head injury?: A case report. *British Journal of Clinical Psychology, 30*, 179-180.

Nebes, R.D., Martin, D.C., & Horn, L.C. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology, 93*, 321-330.

Nelson, H.E. (1982). *National Adult Reading Test*. Winsor, UK: NFER-Nelson.

Nelson, H.E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology, 14,* 259-267.

Nelson, H.E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex, 14,* 234-244.

Nelson, H.E., & Willison, J.R. (1991). *National Adult Reading Test (NART): Test Manual*. *(2nd Ed.).* Winsor, U.K: NFER-Nelson.

Nelson, H.E., & Willison, J.R. (1991). *Supplement to the National Adult Reading Test Manual: Restandarisation of the NART against the WAIS-R,* Windsor, U.K: NFER-Nelson.

O'Carroll, R.E. (1987). The inter-rater reliability of the National Adult Reading Test (NART): A pilot study. *British Journal of Clinical Psychology, 26,* 229-230.

O'Carroll, R.E., Baikie, E.M., & Whittick, J.E. (1987). Does the National Adult Reading Test hold in dementia? *British Journal of Clinical Psychology, 26,* 315-316.

O'Carroll, R.E., & Gilleard, C.J. (1986). Estimation of premorbid intelligence in dementia. *British Journal of Clinical Psychology, 25,* 157-158.

O'Carroll, R.E., Moffoot, A., Ebmeier, K.P., & Goodwin, G.M. (1992). Estimating premorbid intellectual ability in the Alcoholic Koraksoff syndrome. *Psychological Medicine, 22,* 903-909.

Ogden, J.A. (2001). First do no harm. Culturally appropriate neuropsychological assessment for indigenous people: A position paper. *Brain Impairment, 2,* 1-10.

Ogden, J.A., & McFarlane-Nathan, G. (1997). Cultural bias in the neuropsychological assessment of young Māori men. *New Zealand Journal of Psychology, 26,* 2-12.

Patterson, K., Graham, N., & Hodges, J.R. (1994). Reading in dementia of the Alzheimer type: A preserved ability? *Neuropsychology, 8,* 395-407.

Riley, G.A., & Simmonds, L.V. (2003). How robust is performance on the National Adult Reading Test following traumatic brain injury? *British Journal of Clinical Psychology, 42,* 319-328.

Rolleston, R.M. (2001). *The State-Trait Anger Expression Inventory (STAXI): reliability, predictive validity and rewording with a New Zealand (NZ) population.* Unpublished masters thesis. University of Waikato, New Zealand.

Ruddle, H.V., & Bradshaw, C.M. (1982). On the estimation of premorbid intellectual functioning: Validation of Nelson and McKenna's formula and some new normative data. *British Journal of Clinical Psychology, 21,* 159-164.

Ryan, J.J., & Paolo, A.M. (1992). A screening procedure for estimating premorbid intelligence in the elderly. *The Clinical Neuropsychologist, 6,* 53-62.

Sadock, B.J., & Sadock, V.A. (2003*). Kaplan & Sadocks Synopsis of Psychiatry Behavioral Sciences/Clinical Psychiatry.* (4th Ed.). Philadelphia: Lippincott Williams & Wilkins.

Schlosser, D., & Ivison, D. (1989). Assessing memory deterioration with the Wechsler Memory Scale, the National Adult Reading Test and the Schonell Graded Word Reading Test. *Journal of Clinical and Experimental Neuropsychology, 11,* 785-792

Schmand, B., Geerlings, M., Jonker, C., & Lindeboom. J. (1998). Reading ability as an estimator of premorbid intelligence: Does it remain stable in emergent dementia? *Journal of Clinical and Experimental Neuropsychology, 20,* 42-51.

Sharpe, K., & O'Carroll, R.E. (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: A Canadian study. *British Journal of Clinical Psychology, 30,* 381-384.

Silverstein, A. (1987). Accuracy of estimates of premorbid intelligence based on demographic variables. *Journal of Clinical Psychology, 43,* 493-496.

Smith, D., Roberts. S., Brewer, W., & Pantelis, C. (1998). Test-retest reliability of the National Adult Reading Test (NART) as an estimate of premorbid IQ in patients with schizophrenia. *Cognitive Neuropsychiatry, 3,* 71-80.

Stebbins, G.T., Wilson, R.S., Gilley, D.W., Bernard, B.A., & Fox, J.H. (1990). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *Clinical Neuropsychologist, 4,* 18-24.

Taylor, R. (1999). National Adult Reading Test performance in established dementia. *Archives of Gerontology and Geriatrics, 29,* 291-296.

Thompson, N.M. (2005). [A review of the Wechsler Adult reading Test]. In Spies, R. A., & Plake, B. S. (Eds.). *The sixteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology, 24,* 1123-1137.

Vanderploeg, R.D. (1994). *Clinician's guide to neuropsychological assessment*. New Jersey: Lawrence Erlbaum Associates.

Ward, S. (2005). [Review of the Wechsler Adult Reading Test]. In Spies, R. A., & Plake, B. S. (Eds.). *The sixteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Wechsler, D. (1958). *The Measurement and Appraisal of Adult Intelligence* (4th ed.). Baltimore: Williams and Wilkins.

Wechsler, D. (1974). *WISC-R Manual.* New York: Psychological corporation.

Wiens, A.N., Bryan, J.E., & Crossen, J.R. (1993). Estimating WAIS-R FSIQ from the National Adult Reading Test – Revised in normal subjects. *Clinical Neuropsychologist, 7,* 79-84.

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating the WAIS-R from the National Adult Reading Test: A Cross Validation. *Journal of Clinical and Experimental Neuropsychology, 13*, 204-216.

Wilson, R.S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology, 46,* 1554-1555.

Yates, A. (1956). The use of vocabulary in the measurement of intellectual deterioration: A review. *Journal of Mental Science, 102,* 409-440.

Appendix A

Full list of words selected for NZART pilot study.

Caveat	Sieve	Facetious	Catacomb	Chaise
Debris	Ctenophore	Mortgage	Gaoled	Phlegm
Manoeuvre	Maori	Epistle	Thyme	Naïve
Chameleon	Epitome	Risqué	Heir	Equivocal
Eucharist	Colonel	Mousse	Radix	Gauche
Whanau	Cologne	Amygdaloid	Assignate	Touché
Recipe	Chaos	Ochre	Hiatus	Synapse
Torque	Guerrilla	Meringue	Subtle	Whenua
Cheyenne	Tacit	Marquess	Procreate	Tertiary
Epergne	Unique	Hippocrates	Gist	Leviathan
Choir	Ci Devant	Ménage	Gouge	Allele
Gauge	Yacht	Indices	Superfluous	Courteous
Indict	Grotesque	Inadequate	Simile	Sieve
Lingerie	Corps	Caecum	Banal	Campanile
Apophthegum	Subpoena	Chassis	Quadruped	Beatify
Champagne	Talipes	Chord	Cellist	Prelate
Fatigue	Vivace	Ache	Facade	Sidereal
Impugn	Paradigm	Depot	Zealot	Demense
Paroxysm	Tourniquet	Aisle	Drachm	Aver
Crochet	Inertia	Bouquet	Aeon	Labile
Tsar	Legate	Psalm	Placebo	Insatiable
Kaitiaki	Topiary	Capon	Abstemious	Reify
Cognac	Bourgeois	Deny	Detente	Puerperal
Eunuch	Wyvern	Nausea	Idyll	Debt
Reign	Rarefy			

Appendix B

NZART words with pronunciation according to the Phonetic alphabet

Word	Pronunciation	Word	Pronunciation
Debt	det	Reify	`ri:↔,fαI
Choir	`kwαI↔	Cognac	`k□njΘk
Aisle	aIl	Amygdaloid	↔’mIgd↔,l□Id
Chaos	`keI□s	Risqué	`rIskeI
Māori	`mα:□ri	Epitome	↔϶pIt↔mi
Nausea	`n□:zI↔	Indices	`Ind↔,si:z
Grotesque	groY϶tesk	Chassis	□ΣΘsi:
Fatigue	f↔϶ti:g	Superfluous	su:϶p∈:flY↔s
Cologne	k↔ ⇑ loYn	Leviathan	lI`vαI↔T↔n
Subtle	`s℘t↔l	Subpoena	s↔϶pi:n↔
Naïve	nαI϶i:v	Facetious	f↔’si:Σ↔s
Psalm	sα:m	Ochre	`oYk↔
Torque	t□:k	Impugn	Im϶pju:n
Sieve	sIv	Zealot	`zElαt
Whenua	fen϶ua	Façade	fα϶sa:d
Thyme	tαIm	Tourniquet	`tY↔n↔,keI
Lingerie	϶l□)Z↔reI	Hippocrates	hI϶p□kr↔,tI:z
Kaitiaki	kai’tiaki	Quadruped	`kw□drY,pEd
Insatiable	In϶seIΣ↔b↔l	Indict	In϶dαIt
Courteous	`k∈:tIαs	Caveat	⇑ kΘvi:↔t
Hiatus	hαI϶eIt↔s	Corps	k□:
Meringue	m↔’rΘN	Abstemious	↔b϶ sti:mI↔s
Debris	`debri:	Topiary	`t↔YpI↔rI
Inertia	In϶∈:Σ↔	Idyll	`IdIl
Placebo	pl↔϶si:b↔Y	Vivace	v↔϶va:tΣi:
Chameleon	k↔ □mi:li:↔n	Labile	`l↔IbIl
Equivocal	I϶kwIv↔l	Détente	deI϶ta:nt
Crochet	`kroYΣeI	Caecum	⇑ si:k↔m
Tacit	`tΘs↔t	Talipes	`tΘl↔,pi:z
Colonel	`k∈n↔l	Syncope	`sINk↔pI

Appendix C

Information sheet for participants

Hi, my name is TeeJay

I am a Psychology Masters student who is undertaking a thesis that involves writing a psychological test for a New Zealand population. This test is called the National Adult Reading Test or NART for short. The test is used to find out if a person's intelligence has declined after a head injury, disease or mental disorder, and helps psychologists to assess and treat patients.

The problem with the existing NART is that it was normed on a British population and their pronunciation and familiarity with certain words is different from people living in New Zealand and may disadvantage New Zealand subjects by giving them a score that is lower than normal.

My research attempts to find words that are more appropriate to a New Zealand population.

If you choose to take part in this research there are a couple of things you should know.

Firstly this research is entirely confidential, your identity as a participant will not be revealed and any data collected is for the purpose of this research only and will not be given to or seen by persons other than the researcher.

Secondly you as a participant have the right to withdraw from this research at any time and for any reason without any penalty.

Thirdly at the conclusion of this research project the results will be made available to any and all of the participants who wish to obtain it. The results can be obtained by sending an email to tjk12@waikato.ac.nz at which time you will be sent a copy of the results. If there are any queries about the results, I can be reached at the same email address.

Alternatively you may wish to contact my supervisors at the Department of Psychology.

Nicola Starkey and Jo Thakker.

If you have any concerns or questions regarding this research, or regarding confidentially or withdrawing from this research or if you have any questions about the results please ask.

Appendix D

Demographics form

Thank you for taking the time to help me with my research project.

The information collected is entirely confidential and no names will be recorded.

In order to conduct my research I will need certain demographic information to compare to other data I gather.

It would help me greatly if you would complete the following questions, and again I assure you this information remains private and your identity will not be revealed.

Please tick the response that is correct for you

Please state your gender

- o Male
- o Female

Please state your age

- o Under 18
- o 19 – 25
- o 26 –32
- o 33-39
- o 40-46
- o Over 46

Please state your ethnicity

- o Māori
- o European
- o Asian
- o Indian
- o Pacific Islander
- o Don’t know or do not wish to say
- o Other Please State ________________________

Please state your approximate income for the year

- o Under $10,000
- o Between $10,000 and $20,000
- o Between $20,000 and $30,000
- o Over 30,000
- o Do not wish to state or don't know

Please state your highest education level

- o No formal degree
- o High school degree
- o Polytechnic Degree
- o Private training establishment degree
- o Bachelors degree
- o Diploma
- o Honours
- o Masters
- o Postgraduate
- o Other please state ____________________

Is English your first language?

- o Yes
- o No

Do you speak any languages other than English?

- o No Yes (please specify ______________________

Have you ever been treated by a psychiatrist or psychologist?

- o No Yes (Diagnosis ___________________)

Have you ever been treated for drug and/or alcohol abuse/dependence?

- o No Yes (Diagnosis _______________________)

Do you currently have any condition affecting your eyesight or reading ability?

- o No Yes (Diagnosis ________________________)

Appendix E

Test data.

The WASI Full scale, Verbal and Performance IQs predicted from the number of errors made on the NART normed on a New Zealand population.

NART Errors	Predicted Full Scale IQ	Predicted Verbal IQ	Predicted Performance IQ
0	129	128	122
1	128	127	121
2	127	126	121
3	126	125	120
4	125	123	120
5	124	122	119
6	123	121	118
7	122	120	118
8	121	119	117
9	120	118	117
10	118	116	116
11	117	115	115
12	116	114	115
13	115	113	114
14	114	112	114
15	113	111	113
16	112	109	112
17	111	108	112
18	110	107	111
19	109	106	111
20	108	105	110
21	107	104	109
22	106	102	109
23	105	101	108

NART Errors	Predicted Full Scale IQ	Predicted Verbal IQ	Predicted Performance IQ
24	104	100	108
25	103	99	107
26	102	98	106
27	101	97	106
28	100	95	105
29	99	94	105
30	98	93	104
31	97	92	103
32	96	91	103
33	95	90	102
34	94	89	102
35	93	87	101
36	92	86	100
37	91	85	100
38	90	84	99
39	88	83	99
40	87	82	98
41	86	80	97
42	85	79	97
43	84	78	96
44	83	77	96
45	82	76	95
46	81	75	94
47	80	73	94
48	79	72	93
49	78	71	93
50	77	70	92

Appendix F

Test data.

The WASI Full scale, Verbal and Performance IQs predicted from the number of errors made on the NZART normed on a New Zealand population.

NZART Errors	Predicted Full Scale IQ	Predicted Verbal IQ	Predicted Performance IQ
0	124	123	120
1	123	122	119
2	122	121	118
3	121	120	118
4	121	119	117
5	120	118	117
6	119	117	116
7	118	116	116
8	117	115	115
9	116	114	114
10	115	113	114
11	114	112	114
12	113	111	113
13	112	110	113
14	112	109	112
15	111	108	112
16	110	107	111
17	109	106	111
18	108	105	110
19	107	104	109
20	106	103	109
21	105	102	108
22	104	101	108
23	103	99	107
24	103	98	107
25	102	97	106
26	101	96	106
27	100	95	105
28	99	94	105
29	98	93	104

NZART Errors	Predicted Full Scale IQ	Predicted Verbal IQ	Predicted Performance IQ
30	97	92	104
31	96	91	103
32	95	90	102
33	94	89	102
34	93	88	101
35	93	87	101
36	92	86	101
37	91	85	100
38	90	84	99
39	89	83	99
40	88	82	98
41	87	81	98
42	86	80	97
43	85	79	97
44	84	78	96
45	84	77	96
46	83	76	95
47	82	75	94
48	81	74	94
49	80	73	93
50	79	72	93
51	78	71	92
52	77	70	92
53	76	69	91
54	75	68	91
55	75	67	90
56	74	66	90
57	73	65	89
58	72	64	89
59	71	63	88
60	70	62	88