ESTIMATING PREMORBID IQ IN NEW ZEALAND

A Thesis

submitted in fulfilment

of the requirements for the degree

of

Masters of Social Science in Psychology

at the

University of Waikato

by

IRENE TATJANA LICHTWARK

University of Waikato

2011

Abstract

The experience of brain injury changes the world for the person experiencing it and their family. It is important for health providers to know as accurately as possible how severe the brain damage is to be able to deliver the appropriate level of treatment and rehabilitation. Tests are available to measure current cognitive functioning which can be expressed as an intelligence quotient (IQ). One such test is the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV). Other tests are able to estimate premorbid IQ, for example the National Adult Reading Test (NART), the Test of Premorbid Functioning (TOPF) and the New Zealand Adult Reading Test (NZART). The discrepancy between the current IQ and the estimated premorbid IQ scores provides an estimate of the decrease in cognitive function as a result of brain injury. Most of these IQ tests have not been developed or normed for the New Zealand population and their suitability for this population is therefore not known. This study aimed to evaluate the ability of the tests of premorbid IQ to estimate the current WAIS-IV IQ in a New Zealand sample. This sample consisted of 86 New Zealand born, neurologically healthy, men and women (mean age of 46 years), who were administered the WAIS-IV, NART, TOPF and NZART. The results showed that the tests of premorbid IQ significantly over estimated lower IQ scores and significantly under estimated higher IQ scores. New regression formulae for the NART, TOPF and NZART were developed based on the WAIS-IV FSIQ and were found to be only marginally better at predicting current IQ. These new regression formulae also over-and under-estimated current IQ in the lower and upper ranges. The NZART, a New Zealand developed test, showed slightly better performance than the overseas tests. It was concluded that the tests of premorbid

functioning are not very accurate in in their prediction of WAIS-IV current IQ for people in New Zealand and alternative methods of estimating premorbid IQ are suggested.

Acknowledgements

First and foremost I would like to thank my two supervisors, Dr Nicola Starkey and Dr Suzanne Barker-Collo who have supported me throughout this research project with their knowledge and wisdom.

Thank you also to all the participants, you volunteered your time and brain power and I know these tests were not easy to do. Thank you all for your perseverance and generosity.

Special thanks to the University of Waikato for the Masters Research Scholarship which helped to fund this year of study for me.

Finally, I would like to thank my family. Thank you to my parents in Germany, you sparked an insatiable curiosity in me and continue to encourage me to follow my path even if it has led me to the other end of the Earth. Thank you to my husband Colin and our three great children Daniel, Margie and Ida who all had to do without me at times and who patiently and lovingly supported me throughout this journey. This is for you guys.

Contents

List of Tables

List of Figures

Estimating Premorbid IQ in New Zealand.

For a person with brain injury the world has changed. Their neuronal functioning has been altered as a consequence of the injury, and because they are 'on the inside' it might be difficult to understand what has happened. Confusion and insecurity are common consequences (Lezak, 2004). Factual information about these changes is needed for people with brain injury and their families to ease the confusion. An assessment by a neuropsychologist might provide this factual information.

Neuropsychologists assess people with brain injuries to investigate the behavioural consequences of brain damage (Lezak, 2004). They use a variety of methods to gain an understanding of the damage and the current abilities and disabilities of the person with brain injury. They are then able to provide factual information to the patients and their family, as well as recommendations to the professionals about the cognitive effects of the injury, and recommendations or evaluations of treatment plans (Loring & Bauer, 2010).

One of the biggest challenges for the assessing neuropsychologist is that the level of pre injury cognitive functioning is usually not known (Franzen, Burgess, & Smith-Seemiller, 1997). This makes it difficult to determine the extent of the loss of cognitive functioning resulting from the injury. Cognitive functioning is commonly measured as intelligence and this measure provides the intelligence quotient (IQ). The IQ of a person is determined by the use of an intelligence test such as the Wechsler Adult Intelligence Scale (WAIS-IV), for example. Unfortunately, this test only measures the current level of functioning and is unable to access the premorbid (before the injury) level of functioning (Lichtenberger & Kaufman, 2009).

There are tests available which estimate the premorbid IQ by assessing a function that is thought to be resilient to damage by brain injury (Lezak, 2004). These tests provide an IQ score which can be used as an estimated baseline. This estimated premorbid baseline can then be compared to measures of current functioning which provide a comparable IQ score. The relationship between the premorbid IQ and the current functioning IQ allows estimations of the extent of the damage (Lezak, 2004).

The accuracy of some of the tests of premorbid IQ is the focus of this study. Because the construct of intelligence is the basis on which premorbid and current cognitive functioning is assessed this introduction begins with a brief history of this construct. It aims to highlight the arbitrary origins, possibilities and limitations of the construct of intelligence. This is followed by an introduction to and a brief history of the WAIS-IV, a widely used measure of current IQ. An introduction to the methods of estimating premorbid functioning is followed by the description of the tests used in this study to estimate premorbid IQ. An exploration of the assessment of premorbid IQ in New Zealand concludes the introduction.

History of the Construct of Intelligence

> The tendency has always been strong to believe that whatever received a name must be an entity or being, having an independent existence of its own. And if no real entity answering to the name could be found, men did not for that reason suppose that none

> existed, but imagined that it was something particularly abstruse and mysterious (John Stuart Mill, 1806-73 in (Stobart, 2008, p. 139)).

In this quote, the 18th century English philosopher John Stuart Mill may well have been talking about the construct of intelligence. Even now, more than 200 years later the construct of intelligence is still somewhat abstruse and mysterious. One reason for the continuing mystery of the construct of intelligence could be the fact that it is only a construct and not a true biological entity like the senses for example. A construct is defined as an 'invented concept' (Smyth, 2004) and the construct of intelligence is a good example of this. Sadly this is often forgotten (Stobart, 2008). Over time this invented concept has become to be perceived as an actual biological entity which is quantifiable, measurable, and inheritable (Stobart, 2008). The way the western world perceives this construct today is largely a product of influential philosophers' thoughts coupled with different views and belief systems during the last 2500 or so years (Sternberg, 1990). The following brief and incomplete history of the construct of intelligence attempts to demonstrate this.

Mental prowess was highly valued by the ancient Greek philosophers and was seen as distinguishable from other skills (Sternberg, 1990). For example, the abilities to discern and reason where praised by Homer in the 6th century B.C. as especially important qualities in a person which enabled a man to talk well, lead other men, and gain their respect (Homer, 2003).

In the 13th century the monk philosopher Thomas Aquinas observed that some people could not be taught to understand complex topics (Hutchins, 1952c) and he concluded that intellect was stable in a person, and the notion of stability was added to the construct of intelligence.

During the 17th century intelligence became to be seen as a composition of natural and acquired wit (Hutchins, 1952b). Natural wit was gained through experience and use of the mind, whereas acquired wit was seen as a result of formal teaching and cultural influence.

A hierarchy of components of the construct of intelligence was introduced by the German philosopher Immanuel Kant for whom understanding, judgement, and reasoning made up the ‘higher faculties of cognition’ (Hutchins, 1952a). Kant thought that these ‘higher faculties of cognition’ had at least two different forms which distinguished the genius from the masses. The genius would possess creative intelligence, while the masses displayed imitative intelligence. The view that intelligence was stable, quantifiable and hierarchical made the construct measurable (Stobart, 2008).

Finally, Sir Francis Galton perceived intelligence to be hereditary (Rosenthal & Jacobson, 1968). He linked intelligence to physical attributes in people which could be inherited. Galton reasoned that the physical attributes of acute senses would give the brain access to a greater variety of information which would lead to a superior mental ability to reason and judge (Rosenthal & Jacobson, 1968). By the end of the 19th century the concept of intelligence, as perceived in the western world, was strongly focused on mental prowess, was seen as stable, quantifiable, measurable, and inheritable.

It is important to keep the origins of the construct of intelligence in mind when it is used to measure differences in people’s performances on different tasks to avoid misinterpretation (Cohen & Swerdlik, 2005; Wilson, 1999). Because the construct is focusing mainly on mental prowess its measurements tend to neglect a variety of other abilities that humans possess. Or, as Boring (1923) in his often misquoted (Coaley, 2009; Cohen & Swerdlik, 2005; Groth- Marnat, 2009) article states: ‘Intelligence is only what the tests test’ (Boring, 1923, p. 35). In other words, intelligence is only defined by the tests content, while people have a lot more skills and abilities than the western construct of intelligence includes.

With the above in mind, the construct of intelligence is still the most useful construct available today to differentiate between people’s performance. (Boring, 1923; Cohen & Swerdlik, 2005; Groth- Marnat, 2009). Intelligence tests are widely used to predict academic success (Neisser et al., 1996), occupational performance (Hunter, 1986) and serve as detectors of neurological deficits (Lezak, 2004; Loring & Bauer, 2010). The most commonly used intelligence test today is the Wechsler Adult Intelligence Scale, now in its Fourth Edition (WAIS-IV) (Hartman, 2009; Lichtenberger & Kaufman, 2009).

The history of the WAIS-IV began in 1939 with the release of the Wechsler-Bellevue Scale (WB) which was designed to measure verbal and non-verbal skills (Wechsler, 2008). It was divided into 11 subtests and calculated a Full Scale IQ (FSIQ). While the WB was only normed for a small portion of the American population, its successor the WAIS, which was released in 1955, had representative norms which reflected the census data of the entire United States. It also offered scores for Verbal IQ, Performance IQ and FSIQ. The following two versions, the WAIS-R (1981) and the WAIS-III (1997) both retained this scoring structure. Their main differences lay in improvements of the items within the subtests, the addition of two further subtests as well as a greater focus on eliminating errors due to misunderstanding of the tasks by the examinee (Wechsler, 2008). The introduction of the WAIS-IV in 2008 brought several important changes with it. The WAIS-IV and these changes are discussed in detail in the following section.

WAIS-IV

The WAIS-IV is an individually administered battery of tests designed to measure the intelligence of adults aged 16years and 0 month to 90 years and11 month (Wechsler, 2008). It offers five different composite scores as well as a full scale IQ to allow assessment of slightly different aspects of intelligence. These scores have been scaled for 13 age groups and standardised to enable comparison to a mean of 100 and a standard deviation of 15 (Wechsler, 2008). The WAIS-IV is structured into 15 subtests which allow composite scoring of four areas of intelligence, namely Vocabulary Comprehension, Perceptive Reasoning, Working Memory and Processing Speed. Together the scores of these four areas make up the Full Scale IQ. The WAIS-IV is an updated version of the WAIS-III and some of the changes are discussed below.

For the WAIS-IV the subtest structure of the WAIS-III was retained but the composite scores have undergone a major restructuring. In the WAIS-III all subtests were either counted as Verbal Comprehension Index, Working Memory Index, Perceptual Organization Index or Processing Speed Index, which made up the Verbal IQ, Performance IQ and a Full Scale IQ (Cohen & Swerdlik, 2005). In the WAIS-IV these three IQ measures were replaced by four Composite Scores, namely Verbal Comprehension Index Scale (VCI), Perceptual Reasoning Index Scale (PRI), Working Memory Index Scale (WMI) and Processing Speed Index

Scale (PSI). These add up to the Full Scale IQ. A General Ability Index (GAI) can also be calculated from the Verbal Comprehension and Perceptual Reasoning Subscales. It is a faster, easier to obtain measure than FSIQ but clearly less sensitive to working memory and processing speed (Wechsler, 2008). Table 1 displays the four indices of the WAIS-IV and their subtests with a brief explanation of the task for each subtest.

Table 1

The four Indices of the WAIS-IV and their ten Core Subtests with a brief description of the tasks.

Verbal Comprehension Index (VCI)	**Measures 'verbal ability based on reasoning, comprehension and conceptualisation'**.*
Vocabulary	Participant defines words presented orally by examiner. (What does "increase" mean?)
Similarities	Participant describes the relationship between 2 objects/ concepts. ("How are Banana and Apple alike?")
Information	Participant is asked general factual information. ("From which direction does the sun rise?")
Perceptive Reasoning Index (PRI)	**Measure 'nonverbal reasoning and perceptual organisation.'***
Block Design	Working within a time limit the participant must match blocks to geometrical pattern.
Matrix Reasoning	Participant searches for logical patterns in sequences of shapes.
Visual Puzzle	Working within a time limit the participant views a completed puzzle and selects 3 response options, which put together, reconstruct the puzzle.
Working Memory Index (WMI)	**Measures 'specifically, simultaneous and sequential processing, attention and concentration'.****
Arithmetic	Participant performs simple mental arithmetic operations.
Digit Span	Participant repeats a set of digits presented orally.
Processing Speed Index (PSI)	**Measures 'speed of mental and grapho-motor processing.'****
Symbol-Coding	Participant demonstrates visual motor speed and scanning accuracy by transcribing symbols in boxes.
Symbol Search	Participant must search for target symbols in fields of other symbols.

(Table adapted from an unpublished Table by Starkey.)

Note: *quoted from (Wechsler, 2008, p. 9), ** quoted from (Wechsler, 2008, p. 10)

The five supplementary subtests of Figure Weight (PRI), Comprehension (VCI), Cancellation (PSI), Letter-Number Sequencing (WMI), and Picture Completion (PRI) are additional to the core subtests and are not needed to calculate the Composite Scores or FSIQ (Wechsler, 2008).

As stated above the WAIS-IV has many predecessors. These predecessors were deliberately designed without the theoretical base of any particular model of the construct of intelligence in mind. All however, reflect the theory of general intelligence ’g’, which was developed by Spearman in 1927 (Cohen & Swerdlik, 2005). General intelligence ’g’ was seen as the ‘power’ of a person’s intellect, and ‘g’ coupled with varying other factors was responsible for how well a person was able to perform on these tests (Spearman, 1927). Wechsler, the developer of the Wechsler scales, went a step further than his teacher Spearman and perceived these abilities, which made up ‘g’, as different enough to be individually measurable (Cohen & Swerdlik, 2005).

The WAIS-IV still has ‘g’ as an underlying model. Additionally, and for the first time for a WAIS, it is in line with the developers desire to have an intelligence scale which reflects current theories on intelligence, especially those concerned with fluid reasoning, working memory and processing speed. These current theories reflected in the WAIS-IV are discussed in the following paragraphs beginning with the theory of fluid reasoning.

Spearman’s ‘g’ had been split up into fluid and crystalized intelligence to define the various factors of ‘g’ more closely (Neisser, et al., 1996). Fluid intelligence is thought to be genetically determined; it develops throughout the childhood years and becomes fixed from early adulthood onwards. It includes such abilities as reasoning, problem solving and adaptability (Coaley, 2009). Crystallised intelligence is thought to be gained through exercising of fluid intelligence and experience. Because it is dependent on knowledge it peaks later in adult life and is culturally influenced. It includes abilities such as word comprehension and general knowledge (Coaley, 2009).

The choice of subtests in the WAIS-IV reflects the increased focus on fluid and crystallised reasoning. 11 of 15 subtests are thought of as measuring either one of these abilities. The subtests of Block Design, Matrix Reasoning, Visual Puzzles, Figure Weight, Symbol Search, and possibly Coding involve fluid reasoning while Similarities, Vocabulary, Information and Comprehension are measures of crystallised intelligence (Wechsler, 2008).

The second theoretical area that was incorporated into the WAIS-IV was working memory through the introduction of the Working Memory Index (WMI). Working memory is needed to actively keep information in the conscious part of memory where it can be purposefully manipulated and results can be produced (Wechsler, 2008). Good examples of working memory tests are the Arithmetic, Digit Span, and Letter number sequencing subtests in the WAIS-IV. To increase the measure of working memory in the Arithmetic subtest the items were changed to make them mathematically less challenging and easier to understand. This ensured that errors were more likely to stem from working memory deficits than either mathematical or comprehension issues (Wechsler, 2008). To increase the ceiling of working memory testing Digit Span Sequencing was added to the Digit Span subtest (Wechsler, 2008).

The third theoretical concept incorporated into the WAIS-IV was processing speed which led to the introduction of the Processing Speed Index (PSI). Processing speed determines how fast information can be processed by the brain and theoretically processing speed is linked to fluid intelligence. It has been associated with higher performance on cognitive tasks as well as a more effective use of mental resources (Neisser, et al., 1996). Processing speed is sensitive to aging and many neurological disorders for example traumatic brain injury, multiple infarct dementia (Lezak, 2004), or epilepsy (Loring & Bauer, 2010).

There might be some difficulty and confusion about the use of these different new composite scores particularly for experienced users of the WAIS-III. The WAIS-IV Technical and Interpretive Manual (Wechsler, 2008) states quite clearly that,' [t]he terms VCI and PRI should be substituted for the terms VIQ and PIQ in clinical decision-making and other situations where VIQ and PIQ were previously used' (Wechsler, 2008, p. 9). The problem is that the VCI and PRI are not exactly the same measure as the VIQ and PIQ (Loring & Bauer, 2010). The later had subtests measuring working memory and processing speed included in their score while these abilities are now measured separately from the VCI and PRI in the WAIS-IV. A direct replacement of the terms could lead to inaccurate assessments and confusion about which subtests are underlying the composite scores in question. Correlations are high at .89 for VIQ to VIC and .84 for PIQ and PRI (Wechsler, 2008) but still leave room for error. An example of recent research where this direct replacement has been suggested is the study by Barker-Collo, Thomas, Riddick, & de Jager, (2011) on the estimation of premorbid IQ. This study used the WAIS-III and the author quotes the WAIS-IV administration manual to indicate that the difference in terminology was of little consequence.

Another concern stemming from the direct comparison of WAIS-III and WAIS-IV results, which is encouraged by the straight replacement of the terminology, is the possibility of inaccurate judgements about people who have been assessed previously with the WAIS-III. This is also true for the FSIQ as some of the subtests constituting the FSIQ have changed (Loring & Bauer, 2010).

Another factor that complicates a direct comparison of the FSIQs is the Flynn effect. It has been noted that the IQ of a population appears to increase over the years if measured with the same intelligence test (Flynn, 2009). This increase is substantial and in a range of 0.3IQ points per year since the year of norming of the test. In the case of the WAIS-III and WAIS-IV, with 11 years between their norming, people tested at the same time with both tests had a WAIS-III FSIQ that was on average 3.37 points higher than their WAIS-IV score (Flynn, 2009).

Additionally, there is also an issue with the General Ability Index (GAI) and the Full Scale IQ (FSIQ). The GAI is calculated only through the use of the composite scores of VCI and PRI. The manual states clearly that 'the GAI does not replace the FSIQ' (Wechsler, 2008, p. 10) the reason given is that the GAI is less sensitive than the FSIQ due to the exclusion of WMI and PSI from its calculation. So on one hand the direct replacement with a less sensitive measure is prescribed as in the case of VIQ and PIQ, while on the other hand, a little further down a similar replacement is not allowed.

As stated before, these issues are mainly for the experienced WAIS-III user who might be tempted to use only the equivalent to the old VIQ and PIQ measures in their assessment. The correct use of all four composite scores will avoid issues with the composite scores, but during the time of transition between the WAIS-III and the WAIS-IV vigilance is needed when assessment reports are written and read particularly when FSIQs are compared across the two tests.

The theoretical aspects of a test are not its only important feature. It is also very pertinent to have information about the reliability and validity of a test in order to be able to gain a better understanding of the abilities and limitations of a test. Reliability of a test measures how accurate, stable and consistent it is across different situations (Cohen & Swerdlik, 2005). Validity of a test explores the ability of the test to measure a certain construct (Cohen & Swerdlik, 2005). In terms of reliability the WAIS-IV showed good internal consistency ranging from 0.97 to 0.98 across all 13 age groups for the FSIQ and from 0.87 to 0.98 for the factor index scores in the normative sample for the WAIS-IV (Wechsler, 2008). Internal consistency measures the correlation between the test items within a test and their ability to measure the same construct. The internal consistency values obtained by the WAIS-IV are very acceptable. Inter-rater reliability was high and ranged from 0.98 to 0.99 (Wechsler, 2008). Inter-rater reliability measures the correlation between the test results when the test has been administered and scored by different raters. A high inter-rater reliability indicates that the differences in test scores are less likely to be due to the individual administering the test.

In terms of validity, the manual focused on test content, internal structure, correlation with other tests, and special group differences (Canivez & Schraw, 2010). The validity of the test content was shown by the fact that subtests from the same Index correlated higher with each other than subtests from different indices. Internal validity was quoted to be good as all subtests correlated positively to ‘g’. Bowden et al. confirmed these findings in their study of the US and Canadian normative samples for the WAIS-IV and found the internal structure to be very uniform across the two samples (Bowden, Saklfske, & Weiss, 2010). They warn however, about the dangers of incorrect measurements when using the WAIS-IV in countries without norms for this test. For concurrent validity the correlation with the WAIS-III is given as a range from .85 to .94 for the different indices. The comparison of the WAIS-IV with the WISC-IV (Wechsler Intelligence Scale for Children-IV) found a correlation range from .77 to .91. The last group of validity measures was the comparison of WAIS-IV scores of special group samples and demographically matched control groups and as expected, typical results were found (Canivez & Schraw, 2010; Wechsler, 2008).

One purpose of WAIS-IV based assessment reports might be a neuropsychological examination. Such an examination might be performed to gain a baseline measure of a person's current IQ. Knowledge of a person's current IQ can be very useful if the person has experienced a brain injury, because brain injuries are known for their potential to change a person's cognitive functioning, particularly if the injury was severe or diffuse (Lezak, 2004). This current IQ measure, gained during the assessment, can then be compared to measures of premorbid IQ. The bigger the difference is between the premorbid IQ and the current IQ the more severe is the damage to the brain. Knowledge of this severity enables the patient to claim either an appropriate amount of compensation or financial support depending on the legal framework of their country. This knowledge also enables the professionals working with the patient to make recommendations about the intensity, form and duration of rehabilitation programmes (Lezak, 2004). Of course other factors such as location of the injury play a role as well.

Measures of Premorbid IQ

The measures of premorbid IQ, although administered after the injury has happened, are thought to be able to provide an estimate of the person's premorbid IQ. The measures of premorbid IQ are based on three different methods. The first method, the best performance method, bases premorbid IQ estimation on the assumption that individuals perform at a similar level across all areas of functioning and uses the highest score of test results, behavioural observations or historical data as the estimated premorbid functioning level (Lezak, 2004).The second method bases the estimation on demographic data alone (Barona, Reynolds, & Chastain, 1984). The third method relies on current ability as measured in special tests based on the assumption that some abilities are more resilient to brain damage than others (Wechsler, 1958; Yates, 1956) .The present study focuses on the current ability methods so only these will be discussed here. However, before this discussion of current abilities can begin a discourse on the underlying statistical technique related to premorbid IQ estimation is necessary.

Regression. It is important to note that all methods of premorbid IQ estimation rely on a statistical technique called regression. Regression allows the prediction of one variable (dependent variable or outcome) based on the values of another variable (independent variable or predictor) (Field, 2009). This is done by fitting a theoretical line to the data based on the values of the predictor variable(s) from which the prediction points can be calculated. This line can be expressed in the formula y= (a + b x) + error. Where 'y' is the outcome to be predicted, 'a' is the intercept of the straight line fitted to the model and 'b' is the slope of the line while 'x' is the predictor variable and in this case the error score. There is also some residual error as the model will never fit the data perfectly. Two issues connected with the use of regression in the estimation of premorbid IQ need to be discussed here. The first one is about the advantages and disadvantages of using two different types of regression, while the second issue is an inherent problem with the method of regression itself and how to deal with it. It is called ‘regression towards the mean’ and will be discussed further on in this section.

Researchers can choose between two different types of regression namely, linear regression with one predictor and multiple regression with several predictors. Both these types of regression have their advantages and disadvantages which the following discussion aims to highlight.

For ease of discussion, the steps of estimating premorbid IQ with the use of linear regression are described here. The estimation of premorbid IQ involves three steps. The first one is the testing of a sample of neurologically normal participants, which are representative of the underlying population, using a test of intelligence, such as the WAIS-IV, and a test for premorbid IQ estimation. The resulting test scores are related to each other and are used to calculate the regression formulae. The second step is to use the formula with people who have brain injuries of different origins to test if the formula is able to distinguish between different brain injuries. The third step is then to test a person with suspected brain damage with the same tests of intelligence and premorbid estimation and to enter their achieved premorbid estimator scores into the regression formula. If the resulting estimate of premorbid IQ is lower by a predetermined value than the current IQ, as assessed by the application of the WAIS-IV, then it is concluded that some deterioration of IQ has occurred for this person (Lezak, 2004; Veiel & Kooperman, 2001). The above described process of computing a regression formula is based on a linear regression model as only one predictor variable was used.

The use of a single predictor variable in the estimation of premorbid IQ is a tantalising proposition as these tests of premorbid estimation are very quick and easy to administer and easy to score. Historically the Vocabulary subtest of the WAIS-R and later WAIS-III was used, but it was found that single word reading tasks such as the National Adult Reading Test (NART) or the Test of Premorbid Functioning (TOPF) were more resilient to brain damage (Franzen, et al., 1997). Both these tests and how they became to be used as estimators for premorbid IQ are discussed in greater detail below. The suitability of the NART as single estimator of premorbid IQ has been well researched, much more than the TOPF's suitability, and is discussed next.

The NART (Nelson & Willison, 1991) is a single word reading test consisting of 50 words which are irregular in their grapheme to phoneme translation. Crawford, Stewart, Cochrane, Parker, & Besson (1989b), for example, have shown the NART to be a valid measure of intelligence because in their factor analysis they found that the NART loaded highly on 'g' (0.85). In a separate study, Crawford et al. found that the NART alone explained 66%, 72 % and 33% of the variability in FSIQ, VIQ and PIQ respectively (Crawford, Parker, Stewart, Besson, & De Lacey, 1989). Its test scores have proven to be relatively free of the influences of age, gender (Crawford, Parker, & Besson, 1988) and psychiatric diseases such as depression and schizophrenia (Crawford, 1992). Based on these and other studies the NART is seen as a valid measure of premorbid IQ (Berry et al., 1994; Bright, Jaldow, & Kopleman, 2002; Crawford, Deary, Starr, & Whalley, 2001; McGurn et al., 2004). The NART and its various versions is one of the most commonly used single measures to predict premorbid IQ (Franzen, et al., 1997).

However, there are other researchers who feel that the NART and its various versions or the TOPF should not be used as single predicting factors in linear regression formulae for premorbid IQ. They argued that such single word reading tests are purely measures of verbal ability and should therefore not be used to estimate performance or full scale IQ (Gladsjo, Heaton, Palmer, Taylor, & Jeset, 1999; Uttl, 2002). These and other researchers proposed that regression formulae based on multiple regression (more than one predictor) would allow for more accurate estimations of premorbid IQ (Watt & O'Carroll, 1999). These proposed, additional predictors such as age, gender, education and occupation, were of demographic nature.

It had been suggested fairly early on that demographic data are good predictors of premorbid ability as they are unaffected by brain insult, due to their historical nature, and closely related to intelligence test scores (Barona, et al., 1984). Various researchers have shown that demographics are able to predict between 25% (Bright, et al., 2002) and 50% of variability in the outcome variables (Crawford, et al., 1988; Watt & O'Carroll, 1999). Crawford et al. (1999) went on to use the combined demographic data and NART scores to compute regression formulae based on multiple regression. These combined predictor variables explained 73%, 78% and 39% of FSIQ, VIQ and PIQ variance respectively when age, gender, education and occupation were used (Crawford et al., 1989) and 42%, 60% and 25 % of variance for FSIQ, VIQ and PIQ when age, gender, education and socioeconomic status were used (Watt & O'Carroll, 1999).

Comparing the above results of the multiple regressions with the findings from Crawford et al.'s (Crawford, Parker, et al., 1989) linear regressions (66% FSIQ, 72% VIQ and 33% PIQ) shows that the combination of error scores and demographic data results in more variance explained then the use of error scores alone. However, the ability of the linear regression method to explain the variance of the scores lies within the range of the multiple regressions' ability to explain the variance of the data. It is of course generally better to use the most accurate method of prediction available, but there might be circumstances were demographic variables are not available or testing time is very limited and only a single word reading task can be administered. It is reassuring to know that in such cases even the less precise method is still reasonably accurate.

The second issue around the use of the statistical method of regression in the estimation of premorbid IQ is the debate about the phenomenon of the 'regression towards the mean'. This phenomenon happens because of the fact that the *mean* of the true IQ scores is always the same value as the estimated IQ score, as can be seen in the illustration of regression towards the mean in graph A of Figure 1, but the *value* of a true IQ score is regressed to a lower *mean* of the estimated IQ as seen in graph B of Figure 1 (Veiel & Kooperman, 2001). This leads to estimated IQ scores which are closer to the mean than the true IQ scores are.

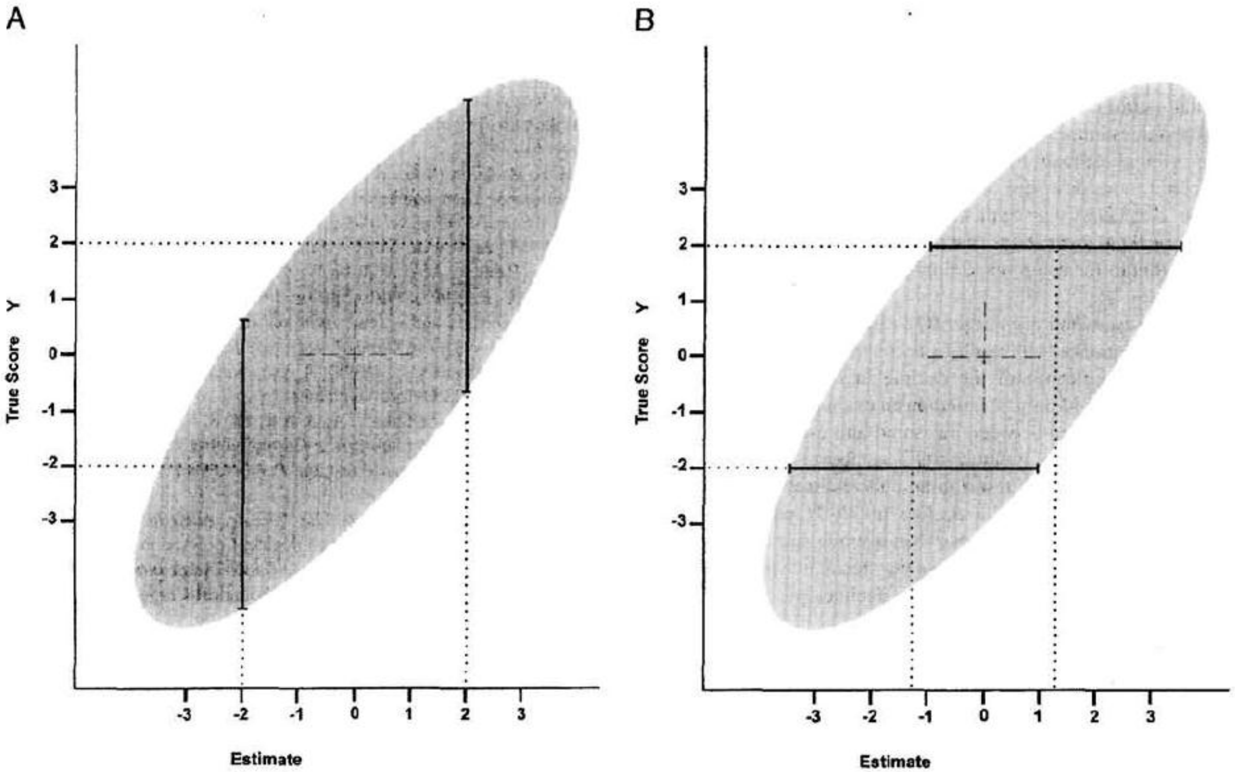


Figure 1. Illustration of regression towards the mean (Veiel & Kooperman, 2001, p. 366).

In practice, this translates to an underestimation of the premorbid IQ for individuals with true IQ scores in the higher regions and overestimation of premorbid IQ for individuals with true IQ scores in the lower regions. More importantly this leads to an underestimation of the damage through brain injury for previously highly functioning individuals and an overestimation of damage for previously lower functioning individuals (Veiel & Kooperman, 2001). There are important implications for these estimation errors depending on the purpose of assessment. Individuals might receive insufficient or unrealistic rehabilitation and in countries were neuropsychological assessments are used in cases of litigation unjust judgement is possible. It is important that practitioners using regression formulae are aware of this phenomenon and view their resulting estimates with the appropriate caution. As so often with statistical techniques regression is not perfect, but it is the most suitable technique for the task at the moment so it will be used in this study as it has been in many others. Now that the issues surrounding regression have been highlighted the discussion can return to different methods of estimating premorbid IQ focusing on the current ability methods.

Current ability methods rely on three different assumptions (Franzen, et al., 1997). The first one is that some cognitive functions are less vulnerable to brain damage than other functions. The cognitive ability to read has been observed to be relatively unaffected by mild to medium grade dementia of the Alzheimer’s type compared to other functions such as memory, reasoning and arithmetic abilities (Lezak, 2004; McGurn, et al., 2004). It has been theorised that reading is a overlearned ability and therefore like a well-worn track ‘etched’ into the brain. These resilient abilities are known as ‘hold’ abilities compared to the ‘don’t hold’ abilities, such as memory and reasoning, which are not so resilient. This distinction has led to the current ability methods being also known as the ‘hold/don’t hold’ methods (Lezak, 2004).

The second assumption of the hold/don’t hold methods is that reading ability is closely correlated with intelligence ‘g’ and therefore a valid measure of intelligence. The subtest with the best correlation with ‘g’ on the WAIS-III and IV is Vocabulary, r_{III} = .88, r_{IV} = .72 (Crawford, Stewart, Cochrane, Parker, & Besson, 1989a; Wechsler, 2008). Furthermore the Vocabulary subtest correlates most highly with level of education which is in itself a good predictor of premorbid functioning (Lezak, 2004). The Vocabulary subtest has traditionally been used to estimate premorbid IQ (Yates, 1956). However, the Vocabulary test requires quite complex responses such as oral definitions of words and is therefore more vulnerable to brain injury than word reading tasks which rely on simpler one word responses (Lezak, 2004). With simpler responses more pure measures can be obtained as fewer cognitive abilities are involved in the response.

The third assumption of this method is that reading irregular words is more resistant to damage that reading regular words. Reading words which are irregular in their grapheme to phoneme decoding relies on previous knowledge of these words and thus minimises the demands on current ability. Reading of regular words, on the other hand, depends on current decoding abilities (Lezak, 2004). These three assumptions of current ability methods have led to the development of several tests to estimate premorbid IQ on the basis of single word reading.

The three word reading tests used in this study were the Test of Premorbid Functioning (TOPF), the National Adult Reading Task (NART), and the New

Zealand Adult Reading Task (NZART). They will be discussed in the following section in the above order.

TOPF

The Test of Premorbid Function (TOPF) is a North American test aiming to provide an estimate of premorbid cognitive functioning in Adults from 20 to 90 years of age (Delis, Kaplan, & Kramer, 2009). It is the new version of the Wechsler's Test of Adult Reading (WTAR) and has been co-normed with the WAIS-IV. The TOPF is made up of a word reading task and several pages of demographic questions. The word reading part of the TOPF (Delis, et al., 2009) is a single word reading test comprised of 70 words which are irregular in their grapheme to phoneme translation and prior knowledge of the words is needed to pronounce them correctly. The words are listed from the easiest ('eye') to the most difficult ('ceilidh') and the participant is required to read them out loud at a comfortable pace. Each correctly pronounced word scores one point, up to a maximum possible score of 70. A phonetic pronunciation guide is provided on the scoring form as well as an auditory guide with which the scorer needs to be thoroughly familiar. Because the TOPF was developed in North America the correct pronunciation is based on North American English.

The TOPF package includes scoring software and offers several scoring models. These include the TOPF Only Model which uses reading data alone, the Simple Demographic model which uses the demographic variables of gender, ethnicity, years of education and occupation to compute the premorbid IQ and the Complex Demographics Predictive Model which uses Demographic, personal factors and developmental factors to provide the estimated premorbid IQ. The software will compute the estimated premorbid IQ and this can then be compared to the current IQ as obtained from the use of the WAIS-IV. However, the option of using the software is only available to scorers of North American samples due to the requirement that the region of the USA is entered in which the participant lives and the fact that the TOPF is normed only for the US population.

The main advantage to be gained from using the TOPF instead of its predecessor the WTAR is that the TOPF is based on the WAIS-IV and not on the WAIS-III as the WTAR. The authors state four revision goals derived from research for the development of the TOPF from the WTAR: increase of the prediction range, improvement of prediction accuracy, expansion of the prediction model and reduction of the effect of brain injury (Delis, et al., 2009). To achieve these revision goals education levels were increased from a maximum of 17 years in education to above doctoral level, higher and lower occupation levels were added and the number of items, especially the more difficult, were increased. Further, some effects of regression towards the mean were eliminated by transforming the TOPF age adjusted scores into WAIS-IV equated scores before entering them into the regression formula. Occupation level and region of the country were added to the regression formulae, personal demographics were added to the test, such as hours of sleep and quality of primary schooling, and the order in which variables are entered into the regression formula has been changed (Delis, et al., 2009).

As the TOPF is a relatively new test in its current form, there are no studies available to gain reliability and validity data from. The TOPF manual, however, stated that the TOPF had a high split half reliability coefficient of r= .92 to r= .99 (M= 0.98, SEM= 2.28) across the age groups, and a good test –

retest stability of *r*=.89 to *r*= .95 (Delis, et al., 2009).Concurrent validity with the WAIS-IV was stated as being r= .37 for PSI and *r*=.75 for VCI.

Because of the lack of research published about the TOPF, research using the WTAR was explored to assess the accuracy and validity of the WTAR and indirectly for the TOPF. Unfortunately a correlation coefficient between the WTAR and the TOPF could not be found in the manual which makes direct comparison of the tests difficult. The WTAR has been well researched in different populations and its test-retest reliability coefficient ranged from *r*= .92 (Thompson & Ward, 2001) to *r*= .97 (Green et al., 2008). Scores from the WTAR related highly with scores from the American NART (*r*= .9) (Thompson & Ward, 2001) and moderately high (r=.66 to .80) with the verbal IQ of the WAIS-III (Thompson & Ward, 2001).

Generally, the reviewed studies found the WTAR a valid measure with good discriminant validity such as Green et al.'s (2008) study for example, but there were some issues with regression towards the mean and underestimation of the premorbid IQ in a healthy Australian sample (Mathias, Bowden, & Barrett-Woodbridge, 2007).

Green et al. (2008) studied the validity of the WTAR as a measure of premorbid IQ in Canadian people with TBI. They administered the WAIS-III subtests of Symbol-Digit-Oral, Similarities and Block Design as measures of current IQ and used the WTAR as well as Vocabulary and Matrix Reasoning from the WAIS-III as measures of premorbid IQ estimates. They tested 25 participants at two and five month post injury and found that there was an improvement in the scores of the measures of current ability which they took as an indicator for recovery. The WTAR performance, however was very stable over the two sessions (M_1= 34.25/50, M_2= 34.21/50, r= .97) which indicated that current performance levels did not influence WTAR performance. The WTAR scores were also compared to premorbid estimated IQ computed with the use of Crawford's demographic regression formulae (Crawford & Allan, 1997) and the WTAR estimates were found to be very similar to the demographic based estimated scores (t(23)=.92, p= .19 for two month and t(23)= 1.076, p= .15) for five month post injury (Green, et al., 2008). Green et al (2008) concluded that the WTAR was a valid measure of premorbid estimation in people with TBI.

A year earlier, researchers in Australia explored the suitability of the WTAR for Australian use (Mathias, et al., 2007). They compared estimated premorbid IQ scores obtained from the administration of the WTAR and the NART to 93 neurologically healthy participants to their current WAIS-III FSIQ and VIQ scores. Two regression formulae were used for the WTAR, one for the UK and one for the US as well as the original British NART formulae. It was found that high IQ levels were underestimated (up to 36 points) while low levels were overestimated (up to 30 points) by all three measures. This was a typical case of regression towards the mean. It was also found that all measures underestimated the premorbid IQ. This study raised two important points. Firstly it showed that word reading tasks cannot be used directly in different English speaking countries as different word familiarity and pronunciation will lead to an underestimation of the non-British or non USA sample. Secondly, it needs to be kept in mind when comparing the WTAR and the NART that they are based on different forms of the WAIS and therefore a direct comparison could lead to distorted results. However, this is an issue that will not go away as long as new test versions are being released. The WTAR has now been developed into the TOPF which has been co-normed with the WAIS-IV while the NART is still normed on the WAIS-R. The use of different regression formulae circumnavigates these issues somewhat, but they should not be ignored because each version of the WAIS measures slightly different abilities and the comparison of different versions across studies increases the error margin of the estimations.

NART

The NART is a single word reading test comprising of 50 words which are irregular in their grapheme to phoneme translation. These words are ordered form the easiest to the most difficult and participants are required to read each word aloud (Nelson & Willison, 1991). The NART was originally developed in 1978 in Great Britain as an assessment tool for the estimation of premorbid IQ in patients with suspected dementia (Nelson & Willison, 1991). It was intended and standardised for the ages from 20 to 70. Subsequently, several studies have shown that the NART can be used with people up to 84 years of age (Nelson & Willison, 1991). The NART was re standardised in 1991 to enable the use of the NART with the WAIS-R, the revised edition of the WB-II (Nelson & Willison, 1991). This revised form of the NART was standardised on 182 neurologically healthy participants. They were assessed with a short version of the WAIS-R consisting of seven subtests. Four of these were verbal subtests. It is not stated why a shortened version of the WAIS-R was used or why the particular subtests were chosen. The WAIS-R results of these participants were used to calculate regression formulae for the NART estimated Full Scale IQ, Verbal IQ and Performance IQ. In either research or clinical work the regression formulae can be used by inserting the obtained NART error score of the examinee into the formulae and comparing the resulting estimations of premorbid IQ with the current WAIS-R IQ obtained from testing the examinee with the WAIS-R.

The manual states that the NART has a high split half reliability (*r*=.93), high inter-rater reliability (r=.96 to *r*= .98) and high test-retest reliability (*r*= .98) (Nelson & Willison, 1991). Sadly there are no details provided about the studies underlying these statements (Calson, 1995).

Criterion validity of the NART was found to be good, with explained variability between 50.2% (Crawford, Stewart, Cochrane, et al., 1989a) and 61% (Crawford, et al., 2001). The authors of the NART quoted a study by Crawford (Crawford, Stewart, Cochrane, et al., 1989a) which stated that the NART loads highly (*r*= .85) on the 'g' factor for intelligence and used this statement to conclude that the NART was a valid assessment tool for intelligence (Nelson & Willison, 1991). Further validation studies are quoted below to demonstrate the validity of the NART in dementia and other neurological conditions.

NART and dementia. A study by McGurn (2004) explored the validity of the use of the NART in patients with dementia. McGurn compared 34 participants with mild to moderate dementia to 464 participants without a diagnosis of dementia. All participants where from the 1932 cohort of the Scottish Mental Survey, where their intellectual abilities had been tested at the age of 11. McGurn found no age related differences on the NART scores between the two groups. The NART estimations of premorbid functioning and the actual scores of functioning at the age of 11 were highly correlated for both groups (r=.63 for the participants with dementia and *r*= 0.60 for those without.) McGurn concluded that the NART was a valid measure for the estimation of premorbid IQ in people with mild to moderate dementia.

The severity of the dementia plays a role in the reliability of the NART as Cockburn, Keene, Hope, and Smith (2003) investigated. They followed 78 participants with confirmed Alzheimer's disease and administered the NART and MMSE (Mini Mental State Examination) annually until each participant's death. Their data analysis was based on assessments over four years. They found that the NART scores decreased as the disease progressed in an irregular pattern compared to the MMSE results. While for some people word reading ability decreased at a similar rate to measures from the MMSE, for others the reading ability was preserved longer or declined a lot sooner than the MMSE measured abilities. It appeared that the decline in reading ability depended on the severity of the dementia and not on age, gender or level of education. Cockburn et al. concluded that although reading ability is not stable in dementia, it deteriorates slower than other cognitive abilities, such as working memory or long term episodic memory, and is therefore a valid measure of premorbid IQ as long as the clinician keeps the severity of the dementia and the resulting danger of underestimating the IQ in more severe cases of dementia in mind (Cockburn & Smith, 2003).

NART and other neurological conditions. Once the NART had been validated for use in people with dementia, its utility for other neurological disorders was explored. Bright, Jaldow and Kopelman (2002) investigated the NART as possibly more accurate than demographic factors in the estimation of premorbid IQ for people with diagnoses of temporal lobe lesions, frontal lobe lesions, Korsakoff's Psychosis and Alzheimer's dementia. Bright et al. compared current IQ measures obtained with the WAIS-III and WAIS-R to NART scores and demographic based scores which had been obtained using Crawford's formula using age, gender, social class and education (Crawford, Stewart, Cochrane, Foulds, et al., 1989). Bright et al. compared 51 neurologically intact participants to 14 participants with Temporal Lobe Lesions, 9 participants with Frontal Lobe Lesions, 35 participants with Korsakoff's psychosis and 32 participants with Alzheimer's Dementia. They found that the NART was not a good estimator of premorbid IQ for people with Korsakoff's psychosis. But in all other conditions the NART's estimations of premorbid IQ where closer to the current IQ scores than the premorbid estimates based on the demographic regression formula. They concluded that the NART should not be used in people with Korsakoff's psychosis but was suitable for the other three neurological conditions (Bright, et al., 2002).

Watt and O'Carroll (1999) investigated the used of the NART in patients with Traumatic Brain Injury (TBI). They compared 25 participants with the diagnosis of TBI to a control group consisting of 50 neurologically healthy people and 20 orthopaedic trauma patients. They used the NART as the estimator of premorbid IQ and the WAIS-R to gain the current IQ of the participants. Watt and O'Carroll found a significant difference (p = 0.01.) in the current IQ scores between the TBI participants ($M_{\text{FSIQ TBI}}$ = 94.70) and the control group ($M_{\text{FSIQ healthy}}$= 107.49, and $M_{\text{FSIQ orthopaedic}}$= 104.88). Furthermore, there was no significant difference (p>.05) in the premorbid IQ estimates between the TBI (M= 24.17 errors) and control groups healthy (M= 22.21 errors) and orthopaedic (M= 22.14 errors) (Watt & O'Carroll, 1999). These results led Watt and O'Carroll to the conclusion that the NART was a valid premorbid IQ estimator for people with TBI.

NART outside Great Britain. The NART had become a versatile and often used research and screening tool in Great Britain. Researchers overseas were becoming interested in this tool quite soon after its appearance. However, a language based test like the NART cannot necessarily be used in different countries without modifications (Franzen, et al., 1997). There were at least three issues that had to be dealt with. First, the regression equation used to estimate the premorbid IQ in Great Britain was likely to be inaccurate in a sample from a different country because the relationship between demographic factors and test scores was likely to be different. Second, people of different cultures, even if they spoke the same language, had different degrees of word familiarity which could change the score on the NART. Lastly, the pronunciation rules which govern the scoring of the NART were based on the English spoken in Great Britain which would greatly disadvantage speakers of other forms of English such as American or Australian for example (Franzen, et al., 1997).

Despite these issues some researchers used the NART without modifying it. The American researchers Ryan and Paolo (1992) explored the suitability of the NART as a procedure to estimate premorbid IQ in the elderly. These researchers assessed 126 neurologically healthy participants with a mean age of 80 with the WAIS-R and the NART. They found high correlations between the NART error scores and WAIS-R VIQ, PIQ and FSIQ of r=-.78,r=-.56 and r= -.74 respectively. These results were similar to the correlations obtained with British samples which range from r= .72 (Lezak, 2004) to r= .81 (Crawford, Parker, et al., 1989). Ryan and Paolo then cross validated their results by administering the NART and the WAIS-R to 20 participants over the age of 75 with neurological impairments. They found that the NART scores were higher than the current IQ scores and concluded that the NART predicted premorbid IQ in their American sample (Ryan & Paolo, 1992).

Sharpe and O’Carroll (1991) likewise used the British NART to explore the utility of the NART in estimating premorbid IQ in a Canadian sample of elderly participants. They compared the NART and WAIS-R scores of 20 elderly people with dementia to 20 elderly people without dementia and computed regression formulae from the results of the healthy participants for FSIQ and VIQ. The NART based estimated FSIQ of the participants with dementia was consistently higher than the NART based estimated VIQ, which in turn was significantly higher than the current WAIS-R FSIQ for the same participants. The researchers concluded that the ability to read words ‘holds’ longer than the abilities underlying the VIQ in dementia (Sharpe & O'Carroll, 1991).

Studies such as these, which use the British NART in countries where the spoken English is different to the British English risk an underestimation of their participant’s premorbid IQ compared to British samples (Franzen, et al., 1997; Lezak, 2004). Therefor other researchers developed various versions of the NART a few of which are discussed below.

NART versions in North America. The North American Adult Reading Test (NAART) was developed in 1989 by Blair and Spreen (1989) for the use with American and Canadian people. These developers retained 35 words of the NART and added a further 26. The resulting NAART was 61 words long and designed to increase the suitability to American and Canadian English as well as ensure a greater word familiarity for people in those two countries (Spreen & Strauss, 1991). A shorter 35 word version, the NAART 35, has also been developed with similar psychometric properties to the NAART (Uttl, 2002) . An additional American version of the British NART was the AMART (American National Adult Reading Test) developed by Grober and Sliwinski in 1991 which consists of 45 words and was intended as an assessment tool for verbal intelligence with early Alzheimer's patients in North America (Franzen, et al., 1997). Further, there exists the ANART (American National Reading Test) which is reportedly more sensitive to small between-individual variations (Gladsjo, et al., 1999; Lezak, 2004) and therefore more appropriate for the use with people of ethnic minorities.

The issues arising from testing ethnic minority groups in general, and with tests developed by ethnic majority groups in particular, are very pertinent to New Zealand (NZ) where the Tangata Whenua (the indigenous people) are now a minority. However, as recently as 2001 researchers ignored the existence of Maori population in their research (Freeman, Godfrey, Harris, & Partridge, 2001). Interestingly, they found that even for relatively recent European immigrants the NART might not be suitable.

The NART in New Zealand. In 2001 Freeman et al. explored the utility of the NART as an estimator of premorbid IQ for people with TBI in NZ. They administered the NART to 65 participants with a history of TBI, 80 participants from the community and 27 orthopaedic patients. Unfortunately the severity of the TBI, or time since the injury are not reported. The participants of the two control groups were neurologically unimpaired (Freeman, et al., 2001). Freeman et.al estimated the premorbid IQ of all participants with the regression formula devised by Crawford for the Scottish population (Crawford, Stewart, Cochrane, Foulds, et al., 1989). This formula utilised the demographic data of age, gender, education and occupation as well as the NART score and calculated the FSIQ, VIQ and PIQ for the WAIS-R. Freeman justified the use of the Scottish formula arguing that most of the participants were of Scottish descent. The estimated IQ, which resulted from the use of the formula, was then compared to the obtained NART scores. An obtained NART score of 11.4 points less than the estimated premorbid IQ was considered as an indicator for impairment (Freeman, et al., 2001). It was found that 30 %, 18 % and 11 % of the participants in the TBI, orthopaedic and community samples respectively had scores that indicated impairment. Freeman et al. concluded that the NART was not a very reliable tool for the estimation of premorbid IQ in people with TBI. The fact that even the participants in the control group had a high rate of impairment was explained as undetected cases with a history of TBI or a possible result of substance abuse (Freeman, et al., 2001).

It is interesting to note that Crawford et al. found only 1% of their 151 neurologically normal participants to be impaired as calculated with a score difference of 15 points (Crawford, Stewart, Cochrane, Foulds, et al., 1989). If Crawford's formula was suitable for the NZ population then a similar percentage would have to be expected for the NZ sample as well. Crawford warned against the use of this formula in other countries as the relationships between IQ and demographic variables cannot be assumed to be the same (Crawford, Stewart, Cochrane, Foulds, et al., 1989). Based on Freeman et al.'s study it can be suspected that Crawford's Scottish formulae are not suitable for the estimation of premorbid IQ in people with TBI in NZ even if the people in Freeman's sample were of the same ethnic origin (Scottish) as those on which Crawford's regression formulae had been based.

The lack of attention paid to the issues of assessing Maori people was addressed in 2003 by Odgen, Cooper and Dudley who were interested in the suitability of neuropsychological tests in NZ in general and in the fairness of these tests for Maori people in particular. They reasoned that tests in general are designed to test what the test developers value and that these values often disadvantage people from cultures other than the test developer's. Maori, the Tangata Whenua of NZ, make up about 15 % of the total population of NZ with cultural roots in the South Pacific. The majority of neuropsychological tests used in NZ are developed in North America and Britain based on the values of the dominant cultures there (Odgen, Cooper, & Dudley, 2003). This can lead to a potential disadvantage for Maori people when tested with common neuropsychological tests. Odgen et al. recruited 20 Maori and 20 Pakeha (non-Maori) participants and administered several commonly used neuropsychological tests to all 40 participants. Some of these tests had been modified to accommodate Maori words or cultural needs. For example seven Maori words were added to the WAIS-R Vocabulary subtest and the Design Fluency Test (DFl) by Jones-Gotman and Miller was included in the test battery as it tests visiospatial abilities which, as the researchers reasoned, are important to Maori culture. In this test participants had five minutes to draw as many designs as they could think of (Odgen, et al., 2003).

Odgen et al. found that Maori performed significantly worse on academic skill based tests than their Pakeha counterparts. For example there was a significant main effect for the Vocabulary subtest ($F(3,36)= 7.88$, $p< .01$) with Pakeha scoring higher than Maori participants. On the other hand, the modified version of the Vocabulary test showed no significant differences between Maori and Pakeha participants. The performance of Maori participants on tests of visiospatial skills was similar to Pakeha performance and for tests which are thought of as less influenced by culture the performance of Maori and Pakeha was also similar. The researchers concluded that this small study indicated the importance of developing tests which are suitable for the NZ population and in particular recognise the differences in culture between Maori and Pakeha (Odgen, et al., 2003).Unfortunately, the authors did not state if there were any significant difference in the years of education attained between Maori and Pakeha participants. Differences here could explain some of the variations in the academic performance between the participants.

Barker-Collo came to a similar conclusion in 2008 after exploring the accuracy of the NART with a non- clinical sample of 89 New Zealanders, 14 of whom were of Maori descent (Barker-Collo et al., 2008). In this study Barker-Collo compared the participant's scores of the WAIS-III, the NART and Spot the Word Test (STW). STW is a test of word recognition where the participants are asked to find the true word in each of 60 word pairs. The other word in each word pair is a made up word without meaning (Baddeley, Hazel, & Nimmo-Smith, 1992). For the Pakeha participants the NART and STW scores correlated highly to the WAIS-III FSIQ scores (r_{NART} =.70, p<0.01, r_{STW}= .70, p<0.01), while for Maori participants there was no significant correlation between NART and WASI-III FSIQ scores. Interestingly, the WAIS-III FSIQ correlated highly with the STW scores for Maori participants (r_{STW} =.91, p>0.01). Despite this high correlation the STW was able to only estimate 52% of the current IQs correctly. Barker-Collo concluded that the NART was particularly unsuitable for people of Maori descent, probably as a result of differing word familiarity, and called for the development of a New Zealand version of the NART or at least for NZ regression formulae (Barker-Collo, et al., 2008). Even though the study's sample size of 14 Maori participants is too small to assume much validity of the study, the findings are in line with those of other researchers, unfortunately also with smallish samples. However this is an important topic and further research is warranted.

Barker-Collo, following on from her research in 2008 and incorporating the above findings, devised NZ regression formulae for the NART in 2011 (Barker-Collo, et al., 2011). They administered subtests of the WAIS-III and the NART to 113 participants, aged between 18 and 84. They compare the WAIS-III and NART scores of the 21 (18%) Maori participants with the scores obtained by the 91 (80%) Pakeha participants. Maori participants' mean FSIQ performance on the WAIS-III subtests was significantly lower than the performance of those of Pakeha descent (M_{Maori}= 100.95, M_{Pakeha}= 116.59, p< 0.01), but there was no significant difference between these groups on the NART scores. The demographic variables of age, gender and years of education all had significant effects on WAIS-III and NART scores. Based on these findings Barker-Collo et al. developed regression formulae for FSIQ, VIQ and PIQ. Scores calculated with these formulae explained 82.1% (*SE* 7.21), 71.9% (*SE* 7.23) and 40.5% (*SE* 9.2) of FSIQ, VIQ and PIQ respectively. These new formulae were better able to predict premorbid IQs in the superior and very superior range, while the old British NART formulae were more efficient at predicting average to high average IQ scores. Barker-Collo et al. concluded that these new NART formulae explained a satisfactory amount of variance and could therefore be used with relative confidence in NZ, but a bigger, more representative sample needed to be studied to validate these formulae further.

NZART

At the same time as Barker-Collo et al were developing the regression formulae for New Zealand, Starkey and Halliday (2011) took up the suggestion of Barker-Collo et al. (2008) and developed a New Zealand version of the NART, the NZART. The NZART comprised of 60 short words which were irregular from the normal encoding rules for the English language in their grapheme –to – phoneme translation. To reflect the word familiarity of New Zealanders better 28 words of the NART had been replaced with words more commonly used in New Zealand, for example 'Meringue' and 'Whenua'. The words were listed in order of increasing difficulty.

Starkey and Halliday also paid attention to Odgen's findings about the need for more culturally suitable tests for Maori people. Starkey and Halliday administered the WASI (a short form of the WAIS-III), the NART and the NZART to 63 participants. They calculated the estimated NART FSIQ, VIQ and PIQ using the original British NART formulae and found that the NART equations explained 42%, 49% and 17% of the variance of FSIQ, VIQ and PIQ respectively. They then developed regression formulae for the NZART and found that these were able to explain 46%, 55 % and 19% of the variance of FSIQ, VIQ and PIQ respectively. These figures were lower than Barker-Collo's regression formulae for the NART had achieved (Barker-Collo, et al., 2011). Part of the reason for that was that Barker-Collo et al. included demographic variables into their formulae. Without the demographics their formulae explained 65%, 45.4%

and 22.8 % of variance in their data for FSIQ, VIQ and PIQ respectively (Barker-Collo, et al., 2008). These values are somewhat closer to the ones obtained by Starkey and Halliday (2011). Another reason for the difference in explained variance could be in the use of different WAIS subtests. While Barker-Collo used subtests of the WASI-III, Starkey and Halliday used the WASI. A study using the same WAIS version for both the NART formulae and the NZART would allow a true comparison of the two methods.

The recent release of the WAIS-IV has offered a great incentive to conduct such a comparative study. This present research will use the WAIS-IV scores as a measure of current IQ. The NART, NZART and TOPF will be administered to the participants and regression formulae will be developed based on the NART, NZART and TOPF for FSIQ and VCI. The aims of this study are to evaluate the suitability of the NART, NZART and TOPF as estimators of premorbid IQ in NZ compared to the WAIS-IV current ability scores. An additional aim is to further validate the NZART.

Method

Participants

The majority of participants for this study were recruited through a poster (Appendix A) on the electronic learning platform of the University of Waikato and the electronic platform of the Maori network at the University. Participants over 40 years of age were mostly recruited through contacting community groups such as exercise classes, gardening clubs and bowling clubs. A few additional participants were gained through word of mouth, or through existing participants volunteering their spouses or relatives. All participants lived either in Hamilton, Cambridge or Tauranga. A smaller group of participants from Auckland was added to the Waikato sample. These were recruited through the University of Auckland. Ethical approval for the study had been sought and granted by the University of Waikato, School of Psychology's Human Ethics Committee.

The inclusion criteria for this study were that the participants had to have been born in New Zealand, have English as their first language and did not have a history of a neurological condition, such as stroke, or traumatic brain injury (TBI).

Seventy-five participants were assessed in the Waikato and thirteen in Auckland by a different experimenter. Five participants from the Waikato sample were later excluded because of mild stroke and a history of TBI. The demographic characteristics of the combined sample are shown in Table 2. The sample was predominately made up of females of Pakeha descent. The male to female ratio was 1: 2.67 with 27.7 % of male participants and 72.7 % of female participants. The Maori to Pakeha ratio was 1: 4.4 with 18.18% of Maori participants and 79.55 Pakeha participants. The male participants had spent slightly more years in formal education than the female participants, but for both genders most participants’ occupations were classified in the skilled labourer category (e.g., trades people and nurses). Female participants had spent more time in New Zealand on average than the male participants and the majority of the sample was either single or married. A substantial portion of the female participants was widowed. Overall there were more right handers than left handers in the sample, but more males were left handed than right handed and only the females had ambidextrous individuals in the sample.

Table 2

Demographic Characteristics of the Sample

Variable	Male	Female	Total
	n= 24	*n*= 64	*N*= 88
Age [mean(*SD*)]	40.92 (23)	48.00 (23.17)	46.07 (23.21)
Min- max	18- 89	16- 90	16- 90
Ethnicity [n (%)]			
Maori	2 (8.3)	14 (21.9)	16 (18.2)
Pakeha	20 (91.7)	50 (78.1)	70 (81.08)
Others	2 (8.3)	0	2 (2.3)
Years in formal education [mean(*SD*)]	15.67 (3.52)	14.27 (3.2)	14.65 (3.34)
Occupation [n (%)]			
Student	4 (16.7)	10 (15.7)	14 (16.0)
Labourer/ clerical	1 (4.2)	7 (11.0)	8 (9.1)
Skilled labourer	13 (54.2)	42 (65.6)	55 (62.5)
Professional	6 (25.0)	5 (7.8)	11 (12.5)
Years in NZ [mean(*SD*)]	38.58 (21.91)	46.96 (23.49)	44.69 (23.26)
Marital stat. [n (%)]			
Single	11 (45.8)	19 (29.7)	30 (34.1)
Married	12 (50.0)	19 (29.7)	31 (35.2)
De Facto	1 (4.2)	11 (17.2)	12 (13.6)
Sep./widowed.	1 (4.2)	15 (23.4)	15 (17.4)
Handed [n (%)]			
Right	17 (70.8)	58 (90.6)	75 (85.2)
Left	7 (29.2)	3 (4.7)	10 (11.4)
Ambidextrous	0	3 (4.7)	3 (3.4)

Participants' ages ranged between of 18 and 90 years, with a mean age of 46.07 (*SD*= 23.21). Figure 2 shows the gender distribution of the participants over the different age groups in five year steps. It can be seen that the age distribution of the sample peaks twice. One peak is within the younger age ranges for 16 to 40 while the other peak is in the older age ranges from 56 to 80 years of age. This is due to the recruiting strategies used. The younger age group comprised mostly of students while the older age group were mostly retired people. The age group of 16 to 20 year olds was the largest age group for males and females followed by the 26 to 30 year old group. For the older age groups the 66 to 70 year old group was the largest followed by the 76 to 80 year old group. Four age groups did not have any male participants in them and two of these were in the older age groups which might reflect the smaller number of males in these age groups in the wider population.

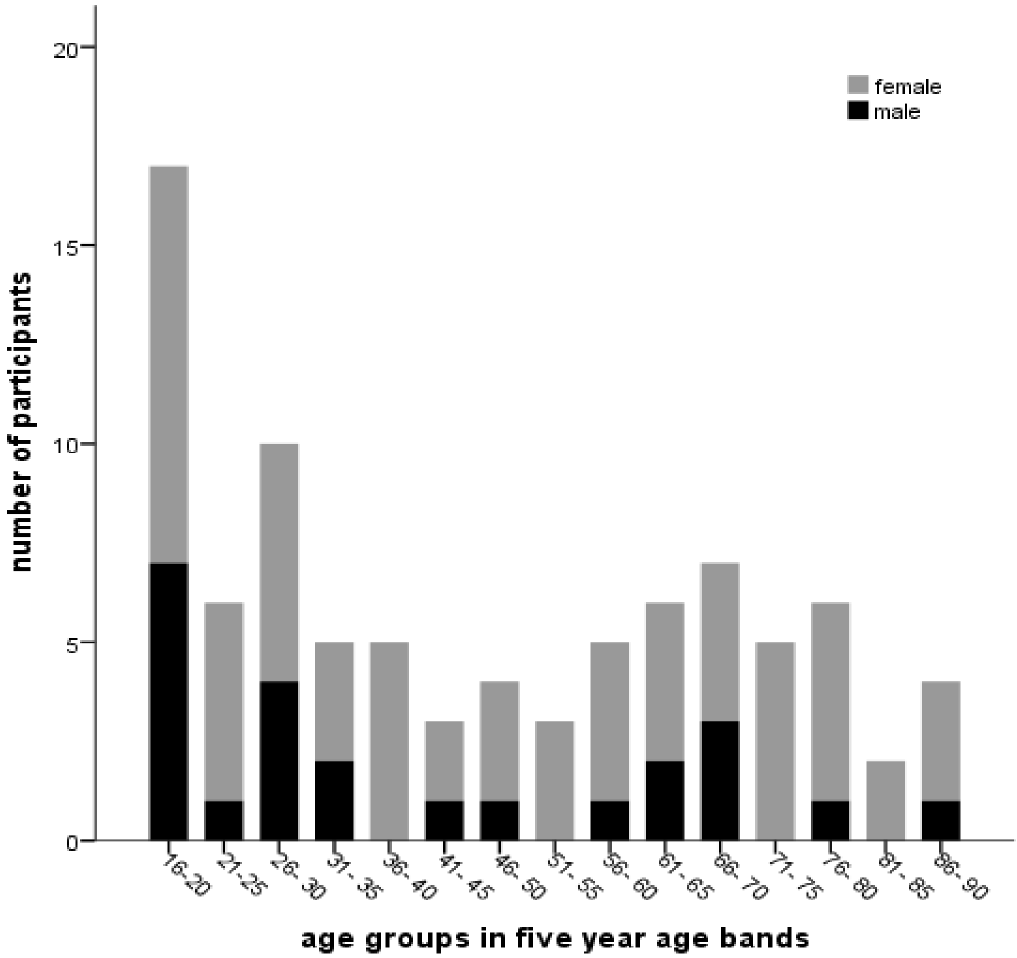


Figure 2. Age distribution of the combined sample in 15 age groups in five year age bands.

Materials

All participants completed the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) to assess current IQ, as well as four tests designed to estimate premorbid IQ. They also completed a short demographic questionnaire which asked about items such as their age, marital status, ethnicity, education and the inclusion criteria mentioned above.

Current IQ measure. The Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) was used as the measure of current IQ. The 10 core subtests of the WAIS-IV were administered in their standard order and procedure as prescribed in the test manual (Wechsler, 2008). The four supplementary subtests were not included as they are not normed for the use with participants over the age of 70 years. Further details of this test have already been presented in the introduction section. The Full Scale IQ (FSIQ) and Verbal Comprehension Index (VCI) were used as the measures of current IQ in this study.

Tests of premorbid IQ. Three tests of premorbid IQ were administered- The Test of Premorbid Function (TOPF) (Delis, et al., 2009), a North American test aiming to provide an estimate of premorbid cognitive functioning in adults from 20 to 90 years of age. It's possible scores range from 0 (no word correct) to 70 (all words correct).

The National Adult Reading Test (NART) is a reading test which was developed in Britain to gain a measure of premorbid cognitive functioning in people who are suspected of suffering from cognitive deterioration (Nelson & Willison, 1991). It's possible error scores range from 0 (all words correct) to 50 (no word correct). For this study, numbers of correct pronunciations were recorded, rather than error scores. to allow a direct comparison with the TOPF scores.

The New Zealand Adult Reading Test (NZART) is the New Zealand version of the NART and was developed as a test of premorbid functioning suitable for the use in New Zealand (Starkey & Halliday, 2011). Again the NZART provides a possible error score from 0 (all words correct) to 60 (no word correct). For the purpose of this study, the NZART scores were recorded as number of words pronounced correctly, to facilitate comparison with the TOPF scores.

All three tests of premorbid IQ were administered separately but in the same way. The participants were given a laminated chart with the words in 2 or 3 columns over both pages printed in large, bold font (font size 20) for each test. They were asked to read the words out loud at their own speed, after the warning that some of the words might be unknown to them or difficult to pronounce. Faulty pronunciation of a word was marked on the score sheet for the NART and NZART while correct reading was awarded 1 point per word for the TOPF scoring sheet. The participant's response was also audio recorded to assist with the scoring later.

Demographics. A demographic form was developed to capture data relevant to this study including age, years in New Zealand, gender, marital status, occupation, ethnicity and cultural identity (see Appendix B).

Because the NART and the NZART share 56% of their words there was the possibility of a learning effect which could result in a higher estimation of IQ from the second measure administered. To minimise the possible learning of the shared words, the two measures were placed as far apart as possible within the administration order of all the tests. To counteract a possible higher IQ estimation due to administration order, the order of administration of these two tests was varied between the different participants. For order 1 the NART was administered as the first measure while the NZART was the penultimate, and for order 2 the NZART was the first measure while the NART was the penultimate one. All other tests were always in the same order: NART (NZART), TOPF, Demographics forms, WAIS-IV, NZART (NART) and STW.

Procedure

Participants recruited at the Universities of Waikato and Auckland were generally enrolled in a first year psychology course at the University of Waikato's Hamilton or Tauranga campuses or the Tamaki campus of the University of Auckland. A particular effort was made to include as many Maori participants as possible; therefor the poster (Appendix A) was added to the electronic platform of the Maori network at University. Participants over 40 years of age were mostly recruited from the community outside University. The researcher contacted and visited various clubs and organisations in the Cambridge and Hamilton area (Bowling Club, Gardening Club, Exercise Classes) to invite participants to the study, and when allowed to do so, presented a brief overview of the study to small groups of people. A list was handed around so interested people could provide their contact details, knowing that the researcher would contact them soon with more information. During this subsequent contact, by telephone, the prospective participant had the opportunity to find out more information about the study, ask questions and to decide if they would like to participate. All 25 contacts that were given were contacted and 23 (92%) agreed to participate.

The order of the test administration was decided by writing alternately either '1' or '2' on the first page of the test sheets as these were printed and put in the appropriate test orders. The researcher then assigned the test orders randomly to the participants.

Participants were assessed in a one to one setting either in an office at the University, or, at the participant's request, in their home. After ensuring that the participants were familiar with the purpose of the study by letting them read the information sheet and answering any questions, the consent form was signed and 2% course credit, or gift vouchers to the value of NZ$ 20 were given. The participants were again reminded of their right to stop participating at any time without having to give an explanation. To ensure the participant's identity was kept confidential in the study, each participant was given a number which was used instead of the name of the participant on all forms. The participant's responses were audio recorded during the administration of the NART, NZART and TOPF for scoring purposes (after gaining permission from the participant). They were assured that only the researcher and maybe one other person would listen to these recordings and that the recordings would be destroyed as soon as the scoring process was finished. The recordings were all kept in a password protected computer, and are only identifiable by the participant's study number.

The tests were then administered as described above. The participants took between 90 to 120 minutes to complete the assessment. At the end of the assessment, participants were thanked for their time and any questions they had about the study were answered.

Each participant's pronunciation during the NART, NZART and TOPF was recorded and listened to later. Special attention was paid to the words which had been marked with a '?' on the scoring forms. The researcher had familiarised herself thoroughly with the pronunciation of each word in the three tests and also had the audio files of the pronunciation guides to listen to. If there was any doubt about the score of a word the audio files were consulted. The researcher listened to the audio files of the pronunciation guides several times during the duration of the data gathering phase to avoid 'examiner drift' (Vanderploeg, 1994) in the scoring during assessments and scoring. The supervisor also listened to 10 % of the recordings to ensure the scoring was accurate. No mistakes were found.

The tests were scored according to the published manuals and the resulting data was entered into PASW. The dependent variables for the current IQ were the WAIS-IV FSIQ and the WAIS-IV VCI scores. The dependent variables for the premorbid IQ were the TOPF scores, and the NART, and NZART error scores which had been reverse scored into a variable called $NART_{correct}$ and $NZART_{correct}$.

Results

The results will be reported in seven sections. The first section gives an overview of the performance of the participants on the different tests and subtests. This is followed by an investigation of the influence of demographic factors on the test scores. The third section reports the comparison of estimated IQ across different measures and the fourth section the comparison of estimated IQ to current WAIS-IV IQ. The fifth section describes the development of the regression formulae for NART and TOPF, while the sixth part of the results section presents the testing of the regression formulae by comparing the resulting premorbid IQ with the obtained WAIS-IV IQ in this sample. The last section contains the exploration of the appropriateness of the word orders in the NART, TOPF, and NZART.

The first step in the data analysis was the recoding of the error scores from the NART and NZART into number correct to facilitate comparison with other test scores and minimise confusion. The error scores of the NART and NZART were only used with existing regression formulae to calculate the estimated IQ for this sample and to calculate the new regression equations.

The NART regression formulae used in this study were the original British formulae found in the new data supplement in the NART manual (Nelson & Willison, 1991). The NZART formulae used were developed by Starkey and Halliday (2011) for a New Zealand sample. The TOPF raw scores were transformed into Scaled Standard Scores with the use of a table in the TOPF manual (Delis, et al., 2009, p. 114). The incidence of missing data was less than 1% and the assumption of homogeneity of the data was met. However, the assumption of normality of the data was violated in the cases of the TOPF and NZART scores as the Kolmogorov-Smirnov test indicated ($D_{\text{TOPF}}(80) = .112$, $p < .05$ and NZART $D_{\text{NZART}}(67) = .109$, $p < 0.05$). Therefore, non-parametric statistics were used for the analysis of these measures.

Participant's Performance

The initial analysis was the description of the participants' performance on the different measures overall and also by gender. Table 3 illustrates the overall performance of the participants on the ten subtests and each index score of the WAIS-IV as well as on the NART, TOPF and NZART. The participant's scores on most of the subtests were just above the mean of 10. This was also the case for the FSIQ and each index where the results were all slightly above the mean of 100 but still within the 'average' category. The TOPF, NART and NZART scores were also slightly above the mean for each test (35, 25 and 30 respectively).

It is of interest to note that the scores on the FSIQ, VCI, WMI and PSI differ significantly between males and females. On all these measures the males scored higher than the females. The effect sizes for these significant results were calculated using the formula for Cohen's ($d = \frac{\mu 1 - \mu 2}{\sigma}$) and the effect sizes were described as suggested by Aron and Aron (2006), with effect sizes of $d \leq .2$ as small, $d \leq .5$ as medium and $d \geq .8$ as large. The effect sizes for the significant results in the above WAIS measures were in the medium range. On the subtests there were significant differences between the genders for Arithmetic and Information. Again the males scored higher than the females and the effect sizes ranged from medium to large.

Table 3

Descriptive Statistics (Mean and Standard Deviation) for the Overall Sample, and , Male and Female Participants' Scores on the Subtests of the WAIS-IV, WAIS-IV FSIQ, VCI, PRI, WMI and PSI as well as TOPF, NART and NZART.

Dependent Variable (scaled)	Overall [mean(*SD*)][a] *N*= 80	male [mean(*SD*)][a] *n*= 59	female [mean(*SD*)][a] *n*= 21	t-test results, [*t*] df= 78	Effect size Cohen's (*d*)
Block Design	12.09 (2.39)	12.42 (2.63)	11.97 (2.30)	.76	.18
Similarities	11.91 (2.82)	12.38 (2.42)	11.74 (2.95)	.89	.23
Digit Span	10.71 (2.61)	11.14 (2.78)	10.56 (2.78)	.88	.22
Matrix Reasoning	11.54 (2.62)	12.29 (2.53)	11.27 (2.63)	1.54	.39
Vocabulary	12.20 (2.53)	12.90 (3.06)	11.95 (2.29)	1.50	.38
Arithmetic	11.30 (2.67)	13.14 (2.56)	10.64 (2.14)	4.01***	.94
Symbol Search	10.98 (2.63)	10.33 (1.80)	11.20 (2.85)	-1.61(*df*=56.28)[d]	-.33
Visual Puzzle	11.40 (3.11)	12.00 (2.76)	11.19 (3.22)	1.03	.26
information	11.03 (2.52)	12.29 (2.51)	10.58 (2.39)	2.78**	.68
Coding	11.29 (2.19)	10.76 (1.55)	11.47 (2.36)	-1.29	-.32
WAIS FSIQ	109 (11.20)	113 (11.54)	108 (6.21)	2.04*	.45
WAIS VCI	108 (13.49)	110 (18.88)	107 (11.05)	2.54*	.22
WAIS PRI	109(12.00)	112 (10.49)	108 (12.34)	1.56	.33
WAIS WMI	105 (13.08)	112 (13.62)	103 (12.21)	2.71**	.69
WAIS PSI	106 (11.10)	103 (5.67)	107 (12.34)	-2.01(*df*=73.06)[d]*	-.36
NART FSIQ	105 (8.55)	106 (7.06)	106 (9.08)	.08	0
NART VIQ	104 (7.86)	105 (6.50)	104 (8.3)	.08	.13
$NART_{correct}$	30 (6.90)	30 (5.70)	30 (7.32)	.079	.02
Nonparametric variables[c]	Overall [median(*IQR*)]	Male [median(*IQR*)]	Female [median(*IQR*)]	Mann-Whitney test [*U*, *z*]	
TOPF FSIQ	109 (18.80)	106 (16.00)	110 (16.00)	577.50,-.46	
NZART FSIQ	109 (14.00)	108 (14.75)	109 (12.50)	329.50, -.64	
NZART VCI	106 (16.00)	105 (17.75)	106 (15.50)	332.00, -.60	
$TOPF_{correct}$	49 (20.75)	45 (21.00)	52 (19.00)	533.00,-.95	
$NZART_{correct}$	43 (15.00)	42 (16.75)	43 (14.50)	332.00,-.60	

p*<.05,*p*<.01,****p*<.001

Note. [a] Subtests scores have mean= 10, *SD*= 3; IQ scores have mean= 100, *SD*= 15.

[b] Scores are not scaled for age.

[c] assumption of normality has been violated

[d] equal variance not assumed

IQR stands for Inter Quartile Range

The next step in the analysis was the exploration of any differences between demographic groups. Possible differences in performance across different ethnicities were considered first. There were two main ethnic groups in the sample: Maori and Pakeha. There were also three participants of other ethnicity but as this sample was very small it was excluded from this part of the analysis. Table 4 summarises Maori and Pakeha participants' mean scores and standard deviations on the different measures and shows the results of the independent sample t-tests performed to explore the significance of the differences between the two groups. Cohen's *d* was again used as a measure of effect size. Because the assumption of normality of distribution of the data for the scores of the TOPF and NZART had been violated the non-parametric Mann-Whitney test was performed to test for differences between these scores.

Table 4 illustrates the overall performance of the participants on the ten subtests, each index score of the WAIS-IV and on the NART, TOPF and NZART. The scores of the participants on most of the subtests were just above the mean of 10 for both Maori and Pakeha. This was also the case for the FSIQ, and each index were the results were all slightly above the mean of 100 but still within the average range. TOPF, NART and NZART average scores were also slightly above the mean for each test (35, 25 and 30 respectively) for both groups. It is also interesting to note that the FSIQ score is the same for both groups. The only statistically significant difference was in relation to Symbol Search where Maori participants scored on average higher than Pakeha participants (medium effect size). This significant result could be due to chance given the high number of comparisons conducted here. Because the sample contained only 16 Maori the finding is most likely not representative for the whole population of Maori.

Table 4

Summary of Maori and Pakeha Participants' Mean Scores and Standard Deviations on the Subtests of the WAIS-IV, WAIS-IV FSIQ, VCI, PRI, WMI and PSI as well as TOPF, NART and NZART.

Dependent Variable (scaled)	Maori [mean(*SD*)][a] *N*= 16	Pakeha [mean(*SD*)][a] *n*= 64	t-test results, [*t*] *df*= 78	Effect size Cohen's [*d*]
Block Design	12.31 (2.21)	12.03 (2.44)	-.42	.12
Similarities	12.88 (3.5)	11.67 (2.60)	-1.54	.43
Digit Span	9.81 (1.68)	10.93 (2.76)	1.56	-.43
Matrix Reasoning	11.12 (2.21)	11.64 (2.72)	.70	-.20
Vocabulary	12.13 (2.28)	12.22 (2.61)	.13	-.04
Arithmetic	11.13 (2.36)	11.34 (2.76)	.29	-.08
Symbol Search	12.38 (2.60)	10.63 (2.54)	-2.45*	.67
Visual Puzzle	11.38 (3.16)	11.41 (3.12)	.04	-0
information	11.63 (2.11)	11.02 (2.63)	-.07	.24
Coding	10.69 (2.63)	11.43 (2.06)	1.06(*df*=19.9)[d]	-.37
WAIS FSIQ	109 (10.94)	109 (11.35)	.03	0
WAIS VCI	110 (8.65)	109 (12.90)	-.41	.08
WAIS PRI	109 (12.00)	109 (12.10)	.10	0
WAIS WMI	103 (9.29)	106 (13.84)	1.00	.23
WAIS PSI	109 (13.10)	105 (10.53)	-1.15	.36
NART FSIQ[b]	103 (8.05)	106 (8.60)	1.45	-.37
NART VIQ	102 (7.40)	105 (7.51)	1.45	-.46
$NART_{correct}$	30.41 (6.93)	27.63 (6.49)	1.45	.40
Nonparametric variables[c]	Maori [median(*IQR*)]	Pakeha [median(*IQR*)]	Mann-Whitney test [*U*, *z*,]	
TOPF FSIQ	102 (15.00)	110 (19.50)	403.00, -1.31	
NZART FSIQ	108 (11.00)[e]	110 (13.75)[e]	330,-.90	
NZART VCI	105 (16.75)[e]	107 (16.75)[e]	328.50, -.93	
$TOPF_{correct}$	43.5(14.50)	52(21.75)	401,-1.34	
$NZART_{correct}$	42.0(12.00)[e]	44(15.75)[e]	328.5,-.93	

Note: $p<.05$
[a] Subtests scores have mean = 10, *SD* = 3; IQ scores have mean= 100, *SD* = 15.
[b] Scores are not scaled for age.
[c] assumption of normality has been violated
[d] equal variance not assumed
[e] n_{Maori}= 15, n_{Pakeha} = 52
IQR stands for Inter Quartile Range

Influence of Demographic Factors

To further investigate the influence of demographic factors on participants' performance across the measures, a series of Pearson's correlations (or Spearman's for non- parametric variables) were conducted. The correlation between the parametric demographic factors (age and years of education) and the WAIS FSIQ, WAIS VCI, NART, TOPF and NZART were investigated using Pearson's correlations. The correlation of occupation to each of the measures was investigated using Spearman's rho. Occupations (based on the best ever job) were categorised into four groups for the analysis and were given ordinal values: 1=Student, 2= labour/ clerical, 3= skilled labour and 4= professional. The correlations between the NART, TOPF and NZART with each other as well as their correlations with the WAIS FSIQ and WAIS VCI were also explored to gain a baseline to which the accuracy of the predicted IQ can be compared.

The demographic factors of gender and ethnicity were excluded from further analysis. The number of males and the number of Maori participants were both too small to allow meaningful correlations. For the same reason these variables were also excluded from the development of the regression equations.

Depression was not investigated despite the fact that 38 % of the participants stated that they had received a diagnosis of depression at some stage in their life. It was not known if the depression was currently occurring or if it was historical and therefor a correlation to the measures of IQ would be meaningless.

Table 5

The Correlations between the NART, TOPF, NZART, Age, Years of Education and Occupation and WAIS FSIQ, WAIS VCI, NART, TOPF and NZART using Pearson's r. Spearman's rho was used for the non-Parametric Variables Occupation. Percentage of Variance Explained is Displayed in Parentheses.

Variables	WAIS FSIQ *N*=80	WAIS VCI *N*=80	NART *N*=80	TOPF *N*=80	NZART *n*=67
$NART_{correct}$	.53**(28)	.57**(32)	-	.73**	.87**
$TOPF_{Std.Score}$	.51**(26)	.42**(17)	.73**	-	.85**
$NZART_{correct}$	.57**(32)	.57**(32)	.87**	.85**	-
Age	.08	.07	.44**	.14	.24*
Years of education	.30**	.38**	.19	.10	.22
Occupation	.081^{s}	.14^{s}	.24*^{s}	.13^{s}	.19^{s}

Note: *Correlation is significant at the 0.05 level

** Correlation is significant at the 0.01 level,

s denotes Spearman's rho correlation

Table 5 summarises the results of the Pearson's correlations as well as the Spearman's rho for the non-parametric variable of occupation. The NART, TOPF and NZART are all significantly and positively correlated with WAIS FSIQ and WAIS VCI. These measures were able to explain between 28% and 32 % of the variability in the scores of WAIS FSIQ and WAIS VCI. As expected no correlations were found between WAIS FSIQ, WAIS VCI and age because these variables had already been scaled for age based on the US norms for the WAIS-IV. A significant positive correlation was found between age and scores on the NART ($p<0.01$), TOPF ($p<0.05$) and NZART ($p<0.05$). This indicated that older participants scored higher on these measures. There was also a significant positive correlation between the years of education and scores on the WAIS FSIQ ($p<0.01$) and WAIS VCI ($p<0.01$) which indicated that higher scores on the measures were related to more years completed in formal education. Occupation was only correlated to the NART and the correlation was positive. This indicates that people with an occupation with a higher ordinal number assigned to it were more likely to have higher $\text{NART}_{\text{correct}}$ scores, or a better job was related to a higher NART score.

Comparison of Different Existing Formulae

Several different regression formulae have been developed for the NART for different samples and based on different version of the WAIS. In this next section of the report the comparison of these formulae in their ability to predict the current WAIS-IV FSIQ and VCI of the present sample is explored. Because previous formulae were developed based on WAIS versions which calculated only the FSIQ, VIQ and PIQ, formulae for the prediction of WMI, PSI and PRI do not exist. Because the NART is a verbal task, only the FSIQ and VCI are investigated here. It is known that the prediction of PIQ (PRI) is less accurate with the NART than FSIQ and VCI (Franzen, et al., 1997).

The three different formulae compared here were the NARTbrit which is the original British formula for the new data supplement taken from the manual of the NART (Nelson & Willison, 1991), the NARTnzS from a paper by Starkey and Halliday (2011) developed for a New Zealand sample and the NARTnzBC also developed in NZ by Barker-Collo Kelly, Riddick, & de Jager (2011). Because the ability to accurately predict current IQ with the NZART and the TOPF is also under investigation in this study, these measures have been included into Table 6. The TOPF IQ estimates are the Standard Scores gained from the TOPF manual and based on a US sample. The equations for the different measures are as follows for **WAIS FSIQ**

NARTbrit FSIQ = 130.6-1.24 x $NART_{error}$

NARTnzS FSIQ = 128.78-1.033 x $NART_{error}$

NARTnzBC FSIQ = 145.716 + (-1.063x$NART_{error}$) + (1.31x education) + (-11.98x ethnic) + (-8.2x gender)

.NZART FSIQ = 124.18-0.903 x $NZART_{error}$

And for **WAIS VCI**

NARTbrit VCI = 127.4-1.14 x $NART_{error}$

NARTnzS VCI = 128.02-1.16 x $NART_{error}$

NARTnzBC VCI = 152.47 + (-1.27 x $NART_{error}$) = + (-0.39 x age) + (1.00 x ethnicity) + (6.92 x gender)

NZART VCI = 123.07- 1.025 x $NZART_{error}$

Table 6 contains the descriptive statistics (mean score and standard deviation) for each formula, and also shows the correlation of each formula with the WAIS-FSIQ and VCI.

It can be seen that all the means for the estimated FSIQ from the formulae are quite close to the mean of the current WAIS FSIQ for this sample. The two mean scores of NARTnzS and the NZART were equally close to the current WAIS IQ score. The NARTnzS formula also had the smallest SD. The NZART had the highest correlation with the scores of the WAIS-FSIQ, while the TOPF had the lowest correlation of all the predictors, which was still significant at the $p<.01$ level. The NARTnzS had the same significant positive correlation as the original British NARTbrit. The positive correlations indicated that participants with a higher estimated FSIQ score were more likely to have a higher current WAIS FSIQ score as well.

A similar pattern is also true for the VCI. Again the TOPF standard Scores had the lowest correlation, which was still significant at the $p<.01$ level but had the closest estimated mean IQ to the WAIS VCI. Once more the NZART had the highest correlation with the WAIS VCI but the calculated mean IQ score was the furthest away for the WAIS VCI mean of scores. All correlations were positive which indicated that participants with higher predicted VCI were likely to score higher on the WAIS VCI as well.

Table 6

Descriptive Statistics of three Different NART Regression Formulae, TOPF and NZART for FSIQ and VCI and their Correlations with the WAIS-IV FSIQ and VCI Respectively.

Measure	WAIS FSIQ Mean(*SD*) *N*= 80	WAIS-IV FSIQ *r*	WAIS VCI Mean (*SD*)	WAIS VCI *r*
NARTbrit	105.61 (8.55)	.53**	104.43 (7.86)	.46**
NARTnzS	107.97 (7.13)	.53**	104.61 (8.02)	.46**
NARTnzBC	114.00(11.42)	.52**	100.72 (11.97)	.47**
NZART FSIQ=	107.19 (9.24)[a]	.57**	103.84 10.58)[a]	.57**
TOPF Standard Scores	106.21(11.90)	.51**	106.21 (11.90)	.42**
WAIS-IV FSIQ	**109 (11.20)**	_	_	_
WAIS-IV VCI	_	_	**108.00 (13.49)**	_

Note. **p<.01,

[a] *n* =67

Comparison of Estimated IQ to Current IQ

An important aspect of a test that estimates IQ is the ability to calculate IQ scores which are reasonably close to the current IQ scores of a person. In fact, in this study where the estimated IQ measures and the current IQ measures were administered at the same time and to the same participants the estimated IQ score should ideally be the same as the current IQ score. To explore how well the different versions of the NART, the TOPF and the NZART were able to predict the current WAIS-IV IQ, the mean of each measure's score was obtained and graphed against the categorised scores of the WAIS-IV. These WAIS-IV categories are a common way for clinicians to describe the level of performance of the examined person in qualitative terms. There are seven categories based on the Scaled Scores obtained (Wechsler, 2008, p. 126):

Extremely low	69 and below
Borderline	70-79
Low Average	80-89
Average	90-109
High Average	110-119
Superior	120-129
Very Superior	130 and above

Figure 3 displays the means of the different measures of estimating IQ categorised into the current IQ categories and compared to the current IQ scores of the WAIS-IV. Only five of the seven categories are displayed because there were no participants in the ‘Extremely Low’ and ‘Borderline’ groups. It can be seen that the IQ estimating measures all overestimate the current IQ score in the ‘Low Average’ and the ‘Average’ categories, and overestimate the current WAIS scores somewhat in the ‘High Average’ category and more so in the ‘Superior’ and ‘Very Superior’ groups.

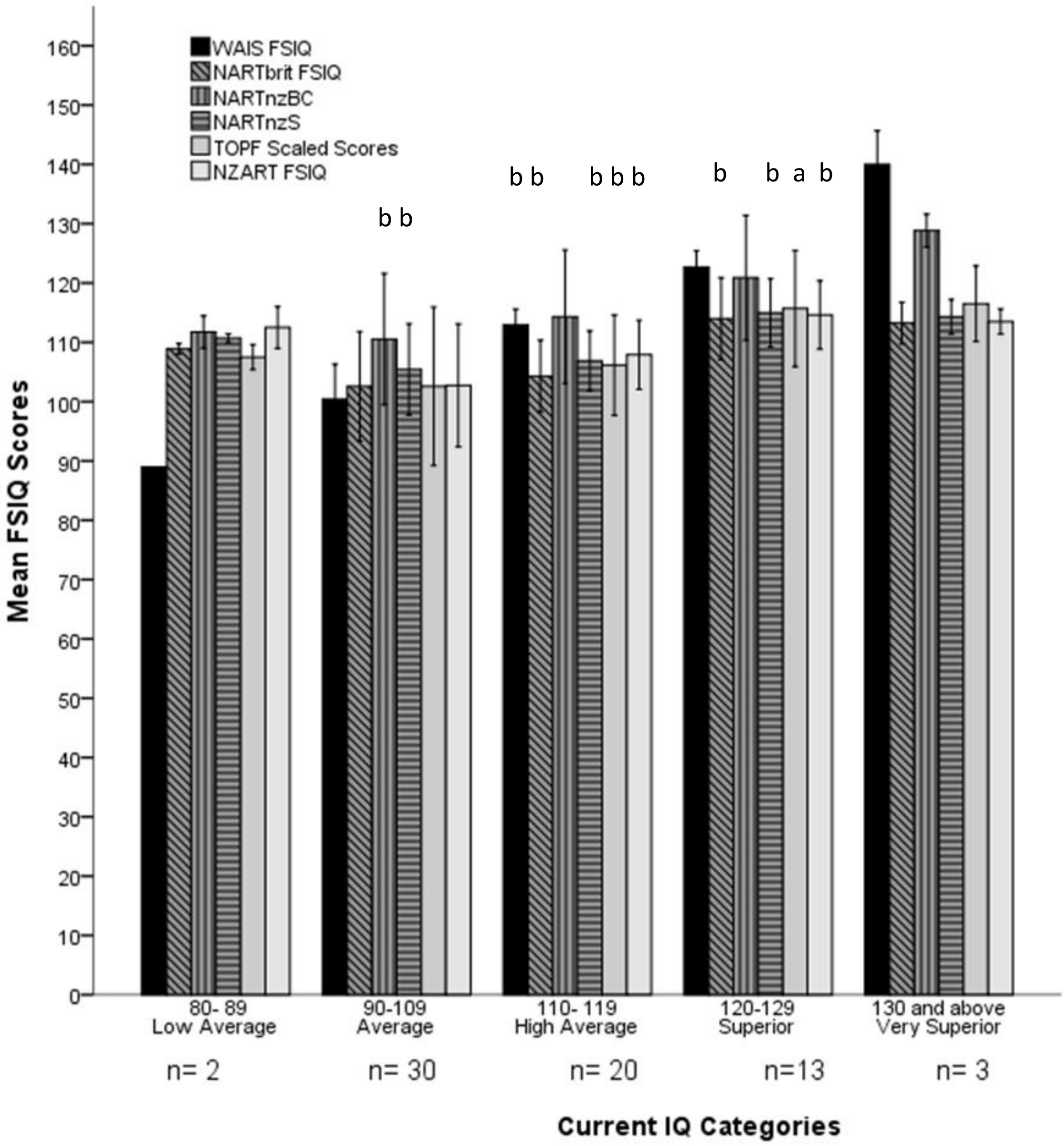


Figure 3. Means and standard deviations of FSIQ scores from the WAIS, NARTbrit, NARTnzS, NZART and TOPF across current IQ categories.

Note. a = p<.05,b = p<.01 compared to current WAIS IQ. Error bar = +/- 1*SD*. N= 68 because the NZART was only administered to the Waikato sample (n = 75) and 8 of these participants had to be excluded.

A repeated measures ANOVA was performed to explore if differences between estimated and current IQ scores were statistically significant in each category. This ANOVA was only conducted for the ‘superior’, ‘high average’ and ‘average’ categories because they each had more than ten participants contributing to the mean score. For some categories tested the Mauchly’s test found that the assumption of sphericity of the data was violated, which is indicated by degrees of freedom with decimal points for the *F* statistics, and the Greenhouse-Geisser estimate of sphericity (ε) was used to correct the degrees of freedom and was reported in those cases. A pairwise comparison was used as a post hoc test with the confidence interval corrected for multiple comparison by the Bonferroni method to avoid Type 1 errors.

The differences between the current WAIS IQ and the estimated IQ were significant in all three groups. For the’ superior’ category (χ^2 (14) = 72.08, $p<$ .001, ε = .44) the results show that there was a significant overall difference between WAIS IQ and estimated IQ scores F (2.20, 26.41) = 5.62, $p<.01$, η^2= .32. The difference between each test’s IQ score and the WAIS IQ score was significant for the NARTnzS (F(1,12) = 21.52), NARTbrit (F(1,12) = 19.67) and NZART (F(1,12) = 27.89) at $p<.01$, for the TOPF (F(1,12) = 6.46) at $p<.05$ and for the NART nzBC (F(1,12) = .37) the difference was not significant (p>.05). The effect sizes were all in the medium range from η^2 = .70 for the NZART to η^2 = .03 for the NARTzBC. The current IQ had been underestimated.

For the ‘high average’ category (χ^2 (14) = 114.88, $p<0.001$, ε= .51) the overall differences between the WAIS IQ and the estimated IQ scores was significant, *F* (2.55, 48.39) = 12.39, $p<.001$with a small effect size of $\eta^2 = .40$. The test of within-subjects contrast showed that only the NARTnzBC mean was not significantly different $F(1,19) = .36$, $p>.05$, $\eta^2 = .02$. All other mean of the estimated IQ scores were significantly different to the WAIS IQ ($p<.001$) ($F_{NARTbrit}(1,19) = 37.90$, $F_{NARTnzS}(1,19) = 25.42$, $F_{TOPF}(1,19) = 10.17$, and $F_{NZART}(1,19) =13.53$)). The effect sizes ranged from a medium $\eta^2 = .67$ for the NARTbrit to a small $\eta^2 = .02$ for the NARTnzBC and the current IQ had been under estimated.

And finally, for the ‘average’ category (χ^2 (14) = 127.55, $p<.001$, ε= .54) the results showed that the overall difference between current WAIS IQ and estimated IQ were significant, $F(2.69,77.98) = 12.41$, $p<.001$. The η^2 effect size of .3 was small. Comparison between each test and the WAIS FSIQ score showed that the NARTnzS ($F(1,29)=17.96$) and NARTnzBC (F(1,29)=35.36) overestimated current IQ significantly ($p<.001$) with a small effect size of .38 and a medium effect size of .55 respectively, while the NARTbrit, the NARTnzBC, the TOPF and the NZART did not have a significant difference to the WAIS IQ scores in this category.

To summarise the comparisons, the estimated IQ scores were overestimating the current WAIS IQ scores in the ‘low average’ and ‘average’ categories, and underestimating the current IQ scores in the ‘high average’, ‘superior’ and ‘very superior’ categories. The differences were found to be significant in the ‘average’, ‘high average’ and ‘superior’ categories across all measures. The significance of the difference could not be determined for the categories of ‘low average’ and ‘very superior’ because the number of participants in each of these categories was below ten.

Finally, a frequency count was performed to determine the number and percentage of scores that were correctly categorised by the NARTbrit, NARTnzS, NARTnzBC, TOPF and NZART. Table 7 displays the results of the frequency count for the existing formulae. It can be seen that all existing formulae predicted the ‘average’ and ‘high average’ categories better than the ‘very superior’, ‘superior’ and ‘low average’ categories. It can also be seen that all formulae except the NARTnzBC predicted the ‘average’ category most accurately. The NARTnzBC was best at predicting the ‘superior’ category.

Table 7

Accuracy of the NARTbrit, NARTnzS, NARTnzBC, TOPF and NZART Formulae for Prediction of WAIS-IV FSIQ Categories.

	WAIS-IV FSIQ				
	Very Superior (n= 3)	Superior (n=14)	High Average (n=24)	Average (n=37)	Low Average (n=2)
NARTbrit					
Very Superior	**0**	-	-	-	-
Superior	-	**4(29)**	-	-	-
High Average	2(67)	6(43)	**4(17)**	6(16)	-
Average	1(33)	4(29)	11(46)	**27(73)**	2(100)
Low Average	-	-	-	4(11)	**0**
NARTnzS					
Very Superior	**0**	-	-	-	-
Superior	-	**4(29)**	-	-	-
High Average	3(100)	7(50)	**10(42)**	6(16)	2(100)
Average	-	3(21)	14(58)	**29(78)**	-
Low Average	-	-	-	1(3)	**0**
NARTnzBC					
Very Superior	**2(67)**	3(21)	2(8)	-	-
Superior	1(33)	**5(36)**	6(25)	7(19)	-
High Average	-	3(21)	**7(29)**	15(41)	1(50)
Average	-	3(21)	9(38)	**13(35)**	1(50)
Low Average	-	-	-	2(5)	**0**
TOPF					
Very Superior	**0**	1(7)	-	-	-
Superior	1(33)	**4(29)**	1(4)	3(8)	-
High Average	2(67)	7(50)	**9(38)**	11(30)	-
Average	-	2(14)	14(58)	**16(43)**	2(100)
Low Average	-	-	-	7(19)	**0**
NZART	(n=2)	(n=13)	(n=20)	(n= 30)	(n=2)
Very Superior	**0**	-	-	-	-
Superior	-	**2(15)**	1(5)	1(3)	-
High Average	2(100)	9(69)	**9(45)**	7(23)	2(100)
Average	-	2(15)	10(50)	**18(60)**	-
Low Average	-	-	-	4(13)	**0**

Note. Numbers in bold represent data which fell in the same category for the prediction and the current FSIQ.

The above formulae to estimate premorbid IQ have all been based on different versions of the WAIS which makes comparison of estimated IQ less accurate. The recent publication of the WAIS-IV offered the opportunity to use this new measure to calculate current IQ and to develop regression formulae for the NART, TOPF and NZART based on the WAIS-IV. Only the TOPF standard scores were developed with the WAIS-IV but regression formula are not given.

Based on these findings, new regression formulae for the FSIQ and VCI were developed for the same sample of participants to explore the possibility of achieving more accurate prediction of current IQ. In the interest of brevity, the development of the formulae related to VCI can be found in Appendix D.

Development of the Regression Formulae

In this fifth part of the results section, the development of the regression formulae for FSIQ is reported. The Regression formulae were generated to predict WAIS-IV FSIQ using the error scores of the NART, TOPF and NZART separately but each combined with demographic factors which had demonstrated a significant correlation with WAIS FSIQ. As noted above, formulae for the WAIS VCI were also developed and are displayed in Appendix D.

The error scores and demographic predictors were entered in a forced entry regression. In forced entry regression the order of the variables is determined by the researcher based on previous research and the known correlations of the predictors to the dependent variable. The highest correlating predictor is entered first, then the second highest and so on. The programme is forced to include all the entered predictors into the different regressions and to calculate each model. Based on the correlation results the demographic factor of education in years was selected as the only predicting variable in the regression equations. Table 8 reports the result of the forced entry regression with WAIS FSIQ as the dependent variable and NART error score and education in years as predictor variables.

As can be seen in Table 8 the NART alone predicted a significant amount of variance in FSIQ (R^2= .28; $F(1,78) = 30.03$, $p<.001$), accounting for 27.8% of the variance in FSIQ. The addition of the demographic factor education significantly improved the model ($R^2_{change} = 0.43$; $F_{change}(1,77) = 4.90$, $p<.05$), accounting for an overall variance in FSIQ of 32.10% with a standard error estimate of 9.35. Therefore, contributing significantly to the prediction were the NART error score ($p<.001$) and education ($p<.05$). The resulting formula for the prediction of FSIQ was:

NART Predicted FSIQ = 114.60 + (-.79 NART error score) + (.74 education in years)

Table 8

Results of the Forced Entry Regression with WAIS FSIQ as the Dependent Variable and NARTerror Scores as Predictor.

	B	*SE B*	*beta*
Step 1			
Constant	126.76	3.33	
NART errors	-.856	.156	-.53***
Step 2			
Constant	114.60	6.39	
NART error	-.79	.16	-.49***
Educ. in years	.74	.34	.21*

Note. R^2 = .28*** for Step 1, delta R^2 = .04* for Step 2,

*p<.05, **p<.01, ***p<.001

The following regression is for the prediction of FSIQ with TOPF scores and years of educations as predictors. Table 9 reports the results from the regression calculations.

The TOPF alone predicted a significant amount of variance in FSIQ (R^2 = 28.6; $F(1,78) = 31.27$, $p<.001$), accounting for 28.6% of the variance in FSIQ. The addition of the demographic factor significantly improved the model (R^2_{change} = .044; $F_{change}(1,77) = 5.12$, $p<.05$), accounting for an overall variance in FSIQ of 33.1% with a standard error of estimate of 9.28. Contributing significantly to the prediction were the TOPF score ($p<.001$) and years of education ($p<.05$). The resulting formula for the prediction of FSIQ was:

TOPF Predicted FSIQ = 77.71+ (.44 TOPF score) + (.75 education in years)

Table 9

Results of the Forced Entry Regression with WAIS FSIQ as the Dependent Variable and TOPF Scores and Education as Predictors.

	B	*SE B*	*beta*
Step 1			
Constant	87.16	4.14	
TOPF score	.48	.90	.54***
Step 2			
Constant	77.71	5.81	
TOPF score	.44	.08	.45***
Educ. in years	.75	.33	.21*

Note. R^2 = .29*** for Step 1, delta R^2 = .04* for Step 2,

*p<.05, **p<.01, ***p<.001

Finally the regression formula for NZART based FSIQ was calculated and Table 10 reports the results of the regression calculation.

The NZART alone predicted a significant amount of variance in FSIQ (R^2 = .32; F(1,65) = 31.14, p<.001), accounting for 32.4% of the variance in FSIQ. The addition of the demographic factor did not significantly improved the model (R^2_{change} = .01; F_{change}(1,64) = 1.80, p>.05), leaving the overall variance in FSIQ at 32.4% with a standard error of estimate of 9.59. Contributing significantly to the prediction was the NZART error score (p<.001). The resulting formula for the prediction of FSIQ was:

NZART Predicted FSIQ = 121.56 + (-.65 NZART error score)

Table 10

Results of the Forced Entry Regression with WAIS FSIQ as the Dependent Variable and NZART Error Scores and Education as Predictors.

	B	*SE B*	*beta*
Step 1			
Constant	121.56	2.50	
NZART errors	-.65	.12	-.57***
Step 2			
Constant	113.67	6.39	
NZART error	-.62	.12	-.54***
Educ. in years	.50	.37	.14

Note. R^2 = .32*** for Step 1, delta R^2 =.02 for Step 2,

*p<.05, **p<.01, ***p<.001

Testing of the new Formulae

The ability of the new regression formulae for the newNART, newTOPF and newNZART to predict FSIQ was tested in the same way as the abilities of the existing formulae had been tested above. To be a good predictor the regression formulae would need to be able to match the prediction quite closely to the actual scores of the WAIS-IV FSIQ of the participants. Of particular interest were the scores at the upper and lower. In order to compare the predicted IQ scores to the current scores the new formulae for each test were used to compute the predicted FSIQ. The resulting scores were then translated into the seven standard current IQ categories listed above and again only five are used because no participants scored in the ‘borderline’ and ‘below average’ categories. The mean score of each category for each test was plotted to allow a visual comparison of the newNART, newTOPF and newNZART mean scores to the WAISFSIQ in each category and the resulting graph is displayed in Figure 4. It can be seen that the estimated IQ scores from the newNART, newTOPF and newNZART are underestimating the current IQ from the WAIS scores in the ‘low average’ and the ‘average’ categories, while the WAIS IQ was overestimated in the ‘high average’, ‘superior’ and ‘very superior’ categories.

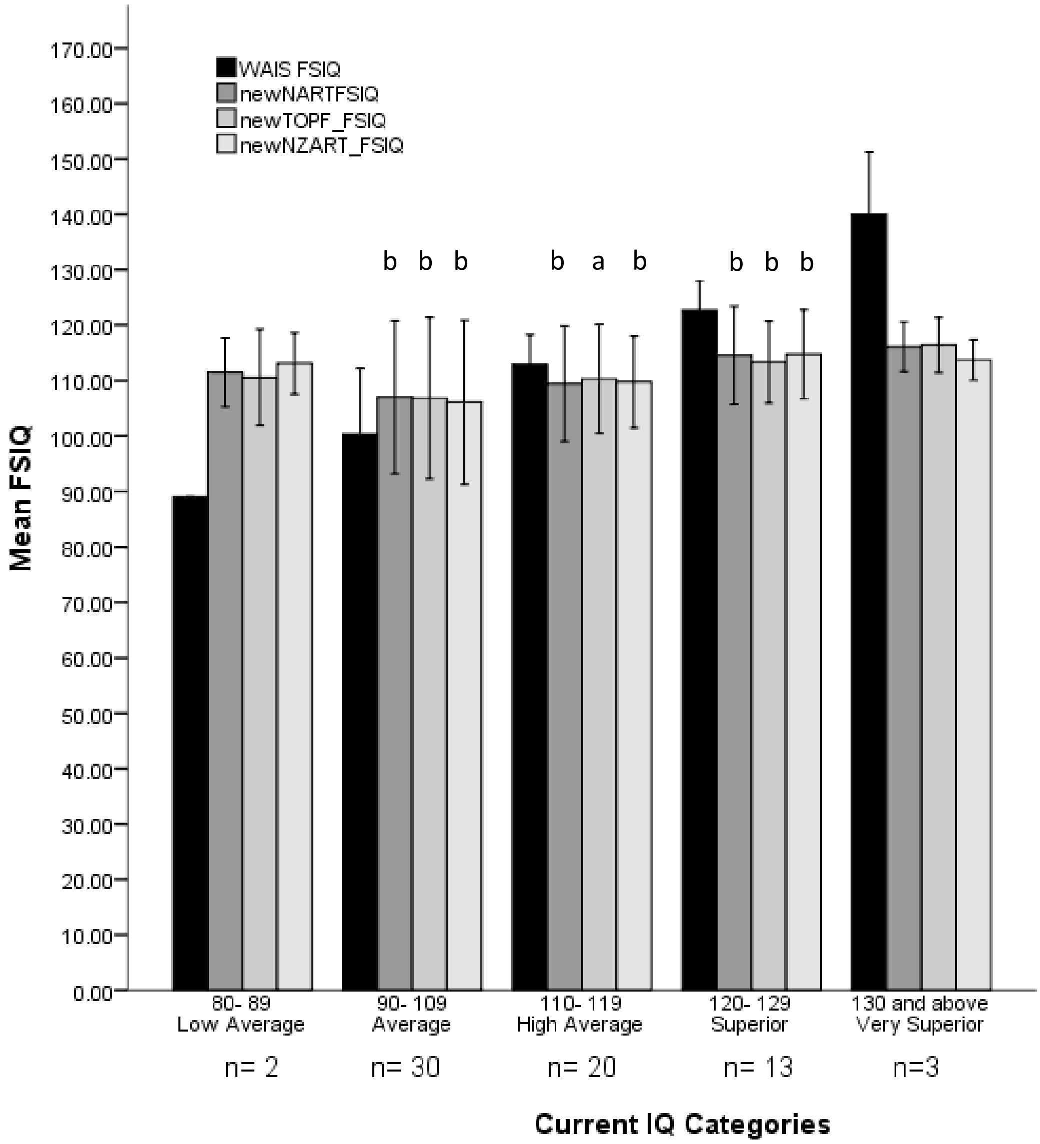


Figure 4. Means and standard deviations of Full Scale IQ from the WAIS-IV, newNART and newTOPF across current IQ categories, $^{a} = p<.05$, $^{b} = p<.01$ compared to current WAIS IQ. Error bar = +/- 1SD.

To explore if the differences between the estimated IQ obtained with the new formulae and the current IQ were statistically significant, a within subject Analysis of Variance (ANOVA) was performed for each category with more than ten contributing scores. There were significant differences between the estimated and the current IQ across all the categories analysed (‘average’, ‘high average’ and ‘superior’). Within-subjects contrasts with Bonferroni corrections were used to explore individual differences.

For the ‘average’ category (n = 30) (χ^2 (5) = 10.59, p<.05) the results show that the difference between estimated and current IQ was significant F(3,87) = 25.81, p<.01, with a small effect size of η^2 = .47. The test of within–subjects contrasts showed that the over estimation of current IQ was significant for each measure ($F_{newNART}$(1,29) = 39.76, $F_{newTOPF}$(1,29) = 39.80,and $F_{newNZART}$(1,29) = 33.26), p<.001 for all measures and the effect sizes of η^2 ranged from .53 for the newNZART to .58 for the newNART and newTOPF.

For the current IQ category of ‘high average’ (n = 20) (χ2 (5) = 8.77, p>.05) the results show that the overall difference between current IQ and predicted IQ are significant, F(3,57) = 6.45, p<.01, with a small effect size of η^2 = .25. Comparison of each estimated IQ from each measure to the WAIS FSIQ showed that the difference was significant for all measures. For newNART (F(1,19) = 11.40) and newNZART(F(1,19) = 9.20) the difference was significant at p<.01, while for the newTOPF(F(1,19) =5.75) the significance was at p<.5. The effect sizes were small and ranges from η^2 = .23 to.38. The current IQ was under estimated.

For the last category analysed,' superior' (n = 13) (χ^2 (5) = 6.15, p>.05) the results show that the difference between current IQ scores and estimated IQ scores were significant, F(3,36) = 27.40, p<.001 and a medium η^2 effect size of .70. Comparison of each estimated IQ to the current IQ show that each measure underestimated the current IQ significantly.(F_{newNART}(1,12) = 32.87, F_{newTOPF}(1,12) = 49.79 and F_{newNZART}(1,12) = 50.55). All differences were significant at p<.001 and the effect sizes of η^2 ranged from a medium .73 to a large .81.

Finally, a frequency count was performed to explore the accuracy of newNART, newTOPF and newNZART for placing the estimated IQ scores into the correct WAIS-IV FSIQ categories. Table 11 shows the WAIS-IV FSIQ categories in comparison to the categories into which the estimated IQ scores have been places. The amount of scores and the percentage of each category are displayed with the number and percentage of correctly categorised scores printed in bold.

It can be seen that the new NART and new TOPF formulae predicted the 'average' category best while the newNZART was better at predicting the 'high average' category. In comparison to the existing NART formulae the newNART predicted the 'average' category somewhat less accurately but was better at predicting the scores that fell in the 'high average' category. The newTOPF predicted both the 'average' and 'high average' scores better than the existing TOPF formula. The newNZART showed a similar performance pattern to the newNART in that it predicted the 'average' categories less well than the existing formulae but was better at predicting the categories for the 'high average' scores.

Table 11

Accuracy of the newNART, newTOPF and newNZART Formulae for Prediction of WAIS-IV FSIQ Categories.

	WAIS-IV FSIQ				
	Very Superior (n = 3)	Superior (n = 14)	High Average (n = 24)	Average (n = 37)	Low Average (n = 2)
newNART					
Very Superior	**0**				
Superior		**2(14)**	1(4)	1(3)	
High Average	3(100)	10(71)	**12(50)**	10(27)	1(50)
Average		2(14)	11(46)	**26(70)**	1(50)
Low Average					**0**
newTOPF					
Very Superior	**0**				
Superior	3(100)	**0**		1(3)	
High Average		12(86)	**10(42)**	11(30)	1(50)
Average		2(14)	14(58)	**25(68)**	1(50)
Low Average					**0**
newNZART	(n = 2)	(n = 13)	(n = 20)	(n = 30)	(n = 2)
Very Superior	**0**				
Superior		**1(8)**			
High Average	2(100)	10(77)	**11(55)**	**10(33)**	2(100)
Average		2(15)	9(45)	20(67)	
Low Average					**0**

Note. Numbers in bold represent data which fell in the same category for the prediction and the current FSIQ.

Exploration of Word Orders

Because the old and new regression formulae of the NART, TOPF and NZART predicted the FSIQ less well than expected each test was examined in more detail. The words on each of these reading tests should be ordered by increasing difficulty. To determine if this is the case for the current sample, graphs were plotted to display the number of correct pronunciations of each word, in order of presentation. If the words were presented in an appropriate order we would expect the graphic depictions of the scores to start with a high number of participants getting the earlier words correct, followed by a steady decrease in the number of correct pronunciations as the words became more difficult.

The number of times each word was pronounced correctly by the participants was counted for each test separately and graphed. For the NART, TOPF and NZART the total correct scores were used where each correct answer scored one point. Figure 5 shows the sum of scores of the participants in the present study for each word of the NART. The words are in their original test order.

Figure 5 shows that the overall trend is as expected. Reading scores are higher on the left side of the graph and lower on the right side. However, several words had a sum of scores which were quite different compared to the words surrounding them. This would indicate that these words appeared to be in the wrong place for this sample of participants. Generally, participants pronounced words with lower scores less often correct than words with higher scores.

To create an order of words which would have suited this sample of New Zealanders better the words were reordered based on their summed correct pronunciations. Figure 6 shows the graph of the reordered NART based on the summed correct pronunciations, with the highest scores on the left and the lowest ones on the right side of the graph. Words number 7, 16, 23, 33, 35, and 42 are all more than 15 positions away from their original position.

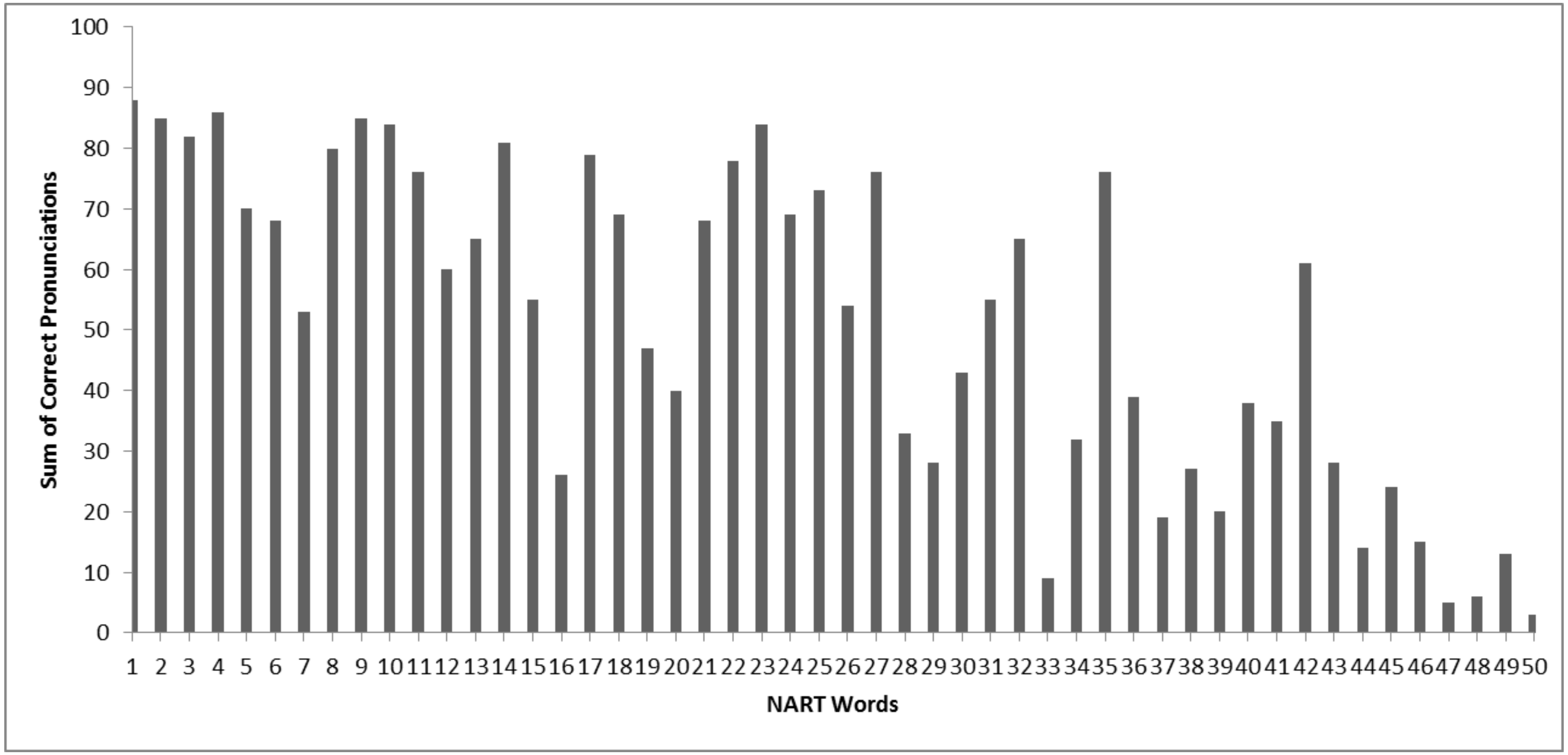


Figure 5. The sum of correct pronunciation for each word of the NART. The words are in their original order.

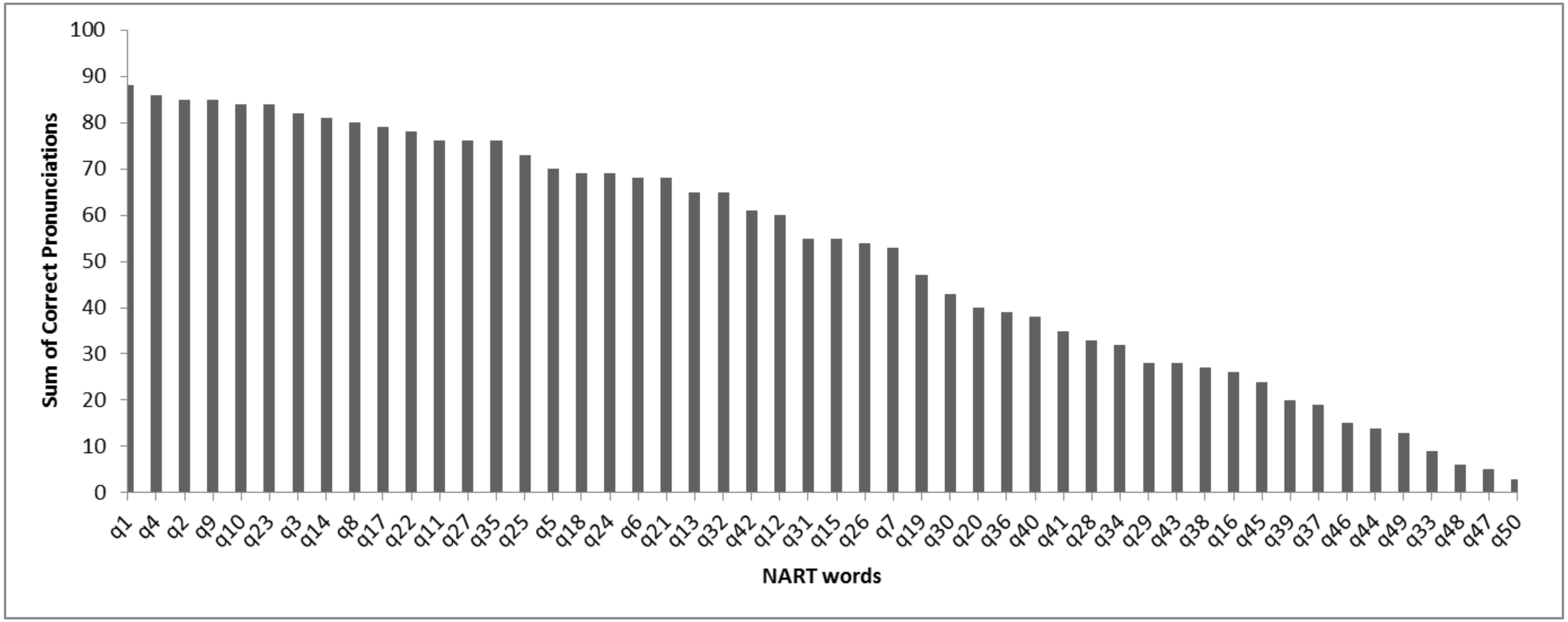


Figure 6. The sum of correct pronunciation for each word of the NART. The words are in their new order based on the sum of correct scores.

The same word order exploration was performed for the TOPF and NZART and the resulting Tables are displayed in Appendix C as they are quite similar to the NART Tables. It was found that for the TOPF only word number 38 was more than 15 positions away from its original position, while for the NZART the words with the numbers 15, 18, 31, and 53 were all more than 15 positions away from their original positions.

To explore if these differences in orders between original word order and order by sums of correct pronunciations were either inherent to our sample or to the measures, the WAIS-IV subtests of Information and Vocabulary were also plotted in the same manner as the other measures. If this difference was inherent to our sample we would expect the WAIS subtests to display a similar difference between the orders as the other measures.

The WAIS-IV subtests are also structured in order of difficulty within each subtest. The easiest question should be at the beginning of the subtest and the most difficult one at the end. This should again lead to a graph which starts with high scores on the left side and finishes with low scores on the right side.

Figure 7 shows the sum of item correct scores achieved by the participants of the study sample for the WAIS-IV subtest of Information.

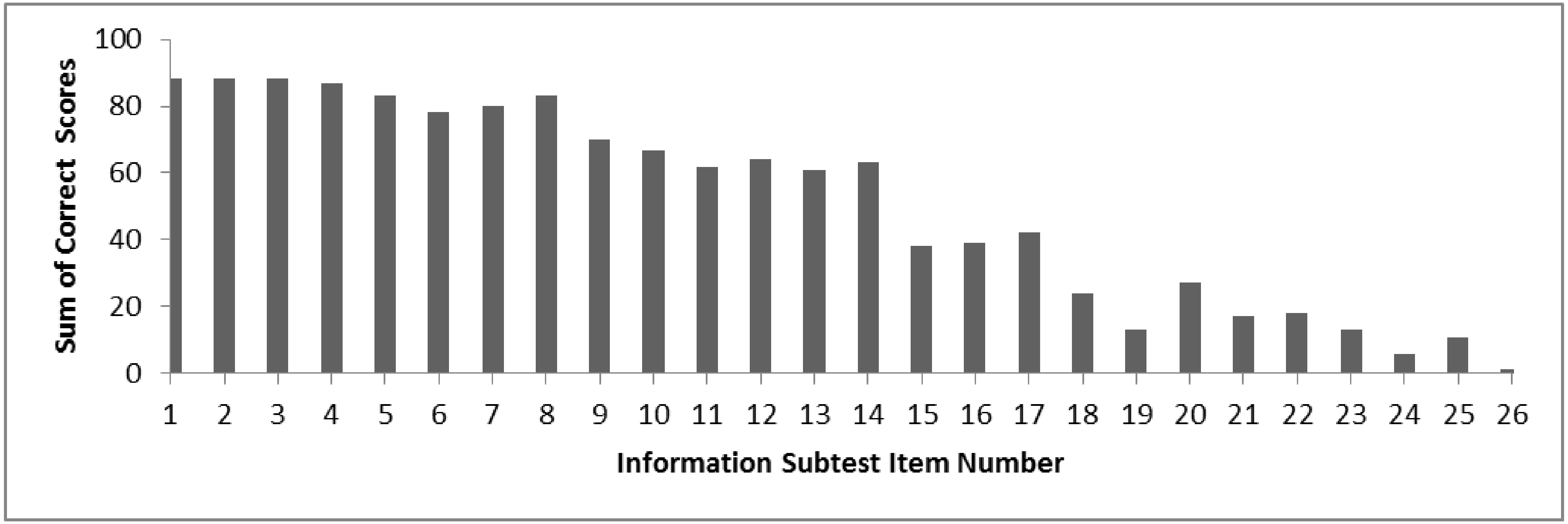


Figure 7. The sum of correct scores for each item on the WAIS-IV subtest of information.

It can be seen in Figure 7 that the sum of items correct scores follows the expected trend quite well. Some variability can be seen and some items with higher or lower scores than their surrounding items might not be in exactly the right position for this sample but overall the graph looks a lot more even than the one of the NART for example in Figure 5.

A further example of WAIS-IV sum of items correct scores subtest plotting is shown in Figure 8. The subtest of Vocabulary has been plotted for sum of items correct scores as well and it can be seen that the overall trend is also as expected. Again, there is some variability in the sums of items correct scores, but overall the items seem to be in approximately the right order for the present sample. The low correct score for the first three items is a result of the scoring procedures. These first items are teaching items and have a maximal score of only one point each while all other items can score up to two points. All other subtests of the WAIS-IV had also been plotted and all followed the expected pattern to a similar extent as the two examples shown here.

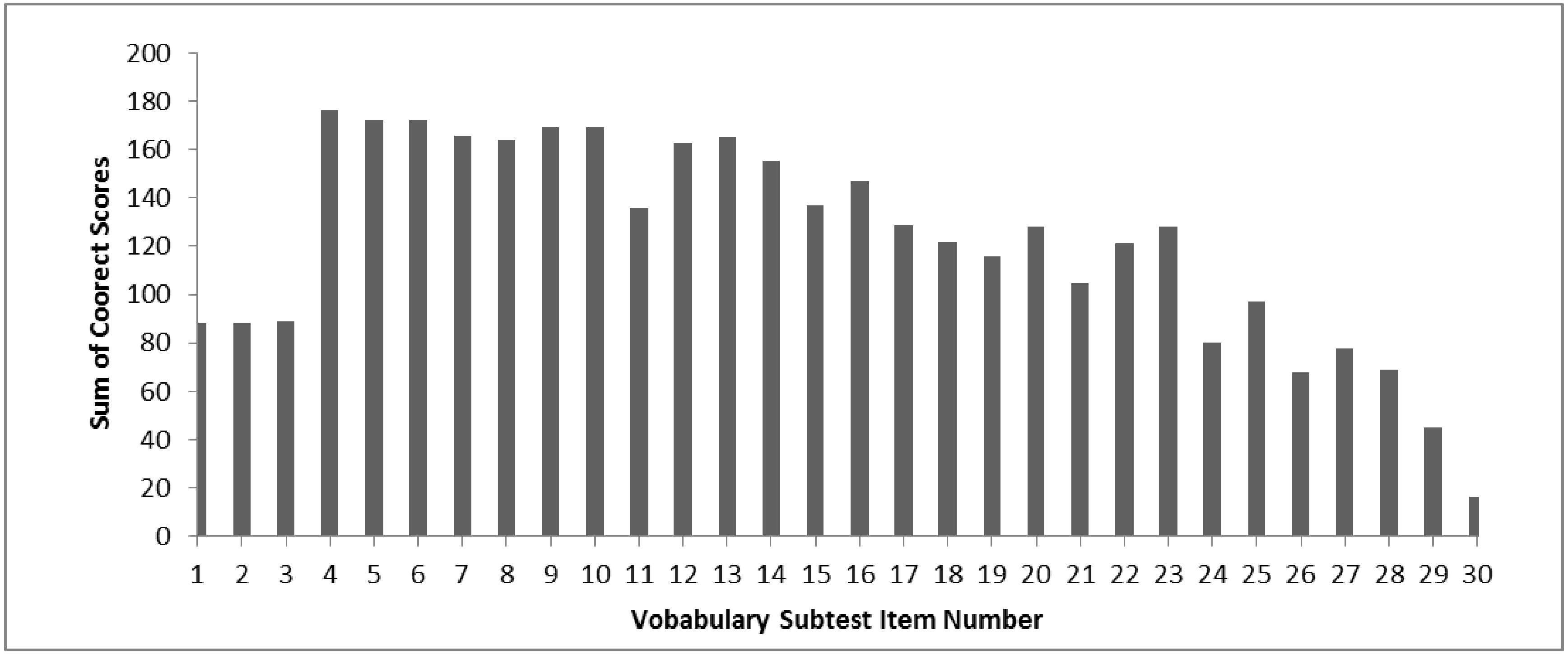


Figure 8. The sum of correct scores of each item for the WAIS-IV subtest of Vocabulary.

Discussion

A person who suffers a brain injury wants to know how their cognitive abilities have changed and what the consequences of these changes are. To answer these questions health professionals treating this person may need to know the level of premorbid cognitive functioning of the person to be able to ascertain how much function has been lost as a result of the injury. Cognitive functioning is often measured as intelligence, using a test of current IQ such as the WAIS-IV. There are also several tests available to estimate the premorbid IQ such as the NART, TOPF and NZART. A comparison of current IQ to estimated premorbid IQ gives information about the extent of the loss of cognitive functioning.

The current and premorbid intelligence tests are normed for populations in different countries namely the United States of America, Great Britain and, in the case of the NZART, for New Zealand. Tests normed overseas may not be appropriate for use in New Zealand, and the only test of premorbid function developed for New Zealand (the NZART) was based on the WASI, rather than the full version of the WAIS.

The present study aimed to evaluate the suitability of the NART, TOPF and NZART as estimators of premorbid IQ in New Zealand compared to the WAIS –IV scores of current IQ. A second aim was to validate the NZART further by testing the previously published regression equations using a different sample.

Summary of Findings

It was found that gender, age, years in education and occupation influenced the scores on the different tests and that ethnicity did not have any influence. The correlations between the NART, TOPF, NZART and the WAIS

FSIQ and VCI were low, with the variance of WAIS FSIQ explained at 28%, 26% and 32% respectively, and for WAIS VCI 32%, 17% and 32% respectively. When the estimated scores were categorised into the IQ categories and compared to the current IQ categories it was found that only the 'average' and 'high average' categories had been estimated with some degree of accuracy which ranged from 35 % to 73% for the FSIQ scores. Current IQ scores in the lower categories were over predicted while higher category IQ scores were under predicted.

The newly developed regression formulae for the NART, TOPF and NZART were able to explain 32%, 33% and 32% of the variance of current WAIS FSIQ respectively and 40%, 32% and 37% respectively of the variance of current WAIS VCI. The estimated IQ scores calculated with the new regression formulae were significantly different to the current IQ scores when categorised into the different IQ categories. Again only the 'average' and 'high average' categories were predicted with some accuracy which ranged from 33% to 70 % for the different categories. And, like the existing formulae the scores in the lower IQ categories were over predicted and the score in the higher IQ categories were under predicted.

The second aim of this study was to validate the NZART further by applying it to a different sample. It was found that the NZART had the highest correlations with the WAIS FSIQ and WAIS VCI of all the measures used to predict current IQ and was able to explain the most variance in current IQ, but there were still statistically significant differences between the current and predicted IQ scores in all categories except for the 'average' range.

In a previous study the variables of age, gender, ethnicity and education all influenced the development of their regression formulae significantly (Barker-Collo, et al., 2011). This was different to our findings where only years of education had a small but a statistically significant influence in the development of the new formulae while the correlations of age, and occupation with the WAIS FSIQ were too small to be included into the regression formulae. For example Barker-Collo et al. found correlations between years of education and WAIS-III FSIQ around $r = .50$ while in our study the correlation between years in education and WAIS-IV FSIQ was as low as $r = .30$ (Barker-Collo, et al., 2011).

Despite gender being also significantly correlated to the WAIS-IV FSIQ we elected not to include it as a variable into the regression formulae because only 27% of the participants were male in our sample and we felt that the inclusion of this variable would not allow valid inferences to the whole population. This resulted in a lower amount of variance explained by the regression formulae.

A surprising result of our study was that ethnicity did not influence the scores at all. This was unexpected because all comparable research had found significant differences in performance between these two groups to some extent (Barker-Collo, et al., 2008; Barker-Collo, et al., 2011; Odgen, et al., 2003; Starkey & Halliday, 2011), and the phenomenon of minority groups performing less well on cognitive tests is well known and documented (Brown, Reynolds, & Whitaker, 1999).

The finding that the accuracy of the categorisations of the scores was relatively low was expected for the overseas developed formulae such as the NARTbrit and the TOPF, but it came as a surprise for the New Zealand developed formulae such as the NARTnzBC, NARTnzS and NZART. Previous research had shown that the prediction of premorbid IQ based on the NARTbrit varied significantly form the current IQ in neurologically healthy participants (Barker-Collo, et al., 2008; Barker-Collo, et al., 2011; Starkey & Halliday, 2011). Additionally, Freeman et al. used the NARTbrit formula to estimate premorbid IQ in a sample with TBI and found results which indicated that the NARTbrit was not a reliable measure of premorbid IQ for New Zealanders with TBI (Freeman, et al., 2001).

It was, however, surprising to find that the NZ developed measures also had significant differences between predicted and current IQ. Theoretically the NZ developed formulae should have performed more accurately, because the relationship between demographic factors and IQ should have been more similar between samples from the same country than between samples taken from different countries (Crawford, Stewart, Cochrane, Foulds, et al., 1989).

Influence of Sample Characteristics on Findings

The demographic structure of the sample for this study was quite different to the demographic structure of comparable studies and this might explain the differing findings somewhat. It is common for research to be conducted with students because they are available to the researcher. However, students are not representative of the whole population of a country. They are generally much younger. The population in New Zealand is ageing and the prediction is that in 2051 half of the population will be over 46 years old compared to a median age of 36.6 years old in the moment (Watkins & Hogenhout, 2010). The mean age of participants in this study was 46 years old, with 47.7% of the participants over the age of 45 years old, and therefore more similar in age to the general population than the samples of the other studies were. These other studies developed the existing NZ formulae with participants of a much lower mean age of 35.38 years (Barker-Collo, et al., 2011) and 25.05 years (Starkey & Halliday, 2011) which makes generalisation for older people problematic. Even though age was used only once as a variable during the development of these regression formulae it would still have indirectly influenced the other variables used, such as years in education.

Years in education was one of the two variables used in the development of the regression formulae and therefore a very influential variable in our study. Because the sample included people from the ages of 18 to 90 years old there was considerable variability in the number of years in education. Particularly the older participants (over 45years of age) in the sample had on average two years less education than the younger participants. Some older participants had to leave school, not because they had reached the limit of their academic abilities, but for economic reasons. It was commented quite regularly by the older participants that they had to stay away from school to earn money or to help with the family and household during the depression in the 1930s and the war. Despite the two years less education the older participants had a slightly higher WAIS FSIQ than the younger participants.

The underlying assumptions around including years in education in the development of regression formulae might well be that people stay in formal education until they have reached their full academic potential. In this way the correlation between current IQ and years in formal education should be large. The fact that the older participants had less years in education but slightly higher

FSIQs may explain why the correlation between years in education and FSIQ was only a small *r* =.30 in this sample. Neisser et al (1996) stated that the correlation between FSIQ and years in education is around *r* = .55 in the US population, and they perceive test scores on intelligence measures as the best predictor for how many years a person is going to spend in education. This US prediction does not factor for economic restrictions to education such as some of the older participants in our study have experienced.

When using years of education as a variable it is important to keep in mind that the *number* of years spend in education is not as important as the *quality* of the education received. It is also important to notice that the older participants had a lot more time to gain informal educational experiences which are not measured here.

The influence of informal education can be seen in the results of this study. Although the older participants had less formal education, they had potentially more exposure to the irregular words used in the tests, which is reflected in the better NART performance of our sample compared to the younger samples of other studies. The participant's mean error score on the NART for our study was 20 (*SD* = 7.37), for the Barker Collo et al. study 31 (*SD* = 6.82) and for the Starkey and Halliday study the mean error score was 26.4 with the *SD* not provided. A further support to this theory is that age was significantly positively correlated with NART scores in this study, which means that people of higher age tended to have higher NART scores as well. Both, informal education and quality of formal education are difficult to measure but can influence the variable of education quite dramatically, and might well be the explanation for the relatively small amount of WAIS FSIQ variance explained by years in education in the regression formula in this sample.

Another finding that was influenced by the characteristics of the sample in our study was the fact that ethnicity did not influence the performance of the participants. As stated above this was an unexpected finding and contrary to existing research which generally found that participants of Maori descent perform less well than their Pakeha counter parts.

Generally this bias is explained with issues in the design, construction, administration and interpretation of the tests which favour the views and abilities valued by the majority groups (Brown, et al., 1999; Odgen, et al., 2003). These issues can lead to members of minority groups generally scoring lower than the participants belonging to the majority group. The 14 Maori participants in this study however, did as well as the Pakeha participants. One reason for this could be the fact that all Maori participants were current students or graduates of the University and as such had a long history of formal education.

A further factor which would have influenced the results of our study, as well as the results of the other cited studies, was the relatively small number of Maori participants in each study. Odgen had 20 Maori participants, BarkerCollo et al. In their 2008 study had 14 Maori participants, and both Barker-Cool et al. and Starkey and Halliday had 21 Maori participants in their 2011 studies. We had 16 Maori participants. These numbers are really too small to base valid conclusions for a population on and most of the authors of these studies acknowledge this, but all still drew conclusions or used these small numbers to base regression formulae on, with ethnicity as a variable in the case of Barker-Collo et al. (2011). To avoid the above criticism we can only conclude that members of minority groups, in this case Maori people, do not necessarily score lower in IQ measures than people of the majority group and that more research with much greater numbers of participants from ethnic minorities needs to be done to allow valid conclusions. It is particularly important to have valid norms for the performance of Maori and other minority groups, because members of minority groups are over represented in brain injury statistics in New Zealand and overseas (Odgen & McFarlane-Nathan, 1997).

Influence of Test Characteristics on Findings

Some of the results may be as a consequence of characteristics which are inherent to the tests used.

In comparing the different NART formulae with the non-NART ones it was found that there was no significant difference between the measures in their ability to predict current IQ. This could indicate that the lack of accuracy in predicting current IQ for this sample is not an issue inherent to the NART alone, but to all the measures used here which rely on the irregular word reading method. In order to investigate why the measures were relatively poor at predicting current IQ, each test was examined in more detail. It was found that the words were not always in their correct order within each test. Ideally the easiest words (as indicated by the highest sum of correct scores across all participants) should be at the beginning of the test and the most difficult words (as indicated by the lowest sum of scores across all participants) should be towards the end of the list of words. A reordering of the words to suit the New Zealand word use better might increase the accuracy of these word reading tests.

As far as we are aware no research has been done on the word orders of these tests. It would be interesting to study the relationship between word order and perceived difficulty of these tests. The early appearance of words which are perceived as difficult could influence how the entire test is seen by the participants, and this perception might influence their scoring on these tests. Some New Zealand research has suggested that the NART was not very well liked by their participants and was perceived as difficult to complete (Odgen, et al., 2003). The NART in particular is an often used measure to estimate premorbid IQ after brain injury in New Zealand and overseas, and it is important for this user group to have a test which is perceived as not too difficult, tiring or unpleasant (Lezak, 2004).

Another test characteristic investigated was the fact that neither the existing nor the newly developed regression formulae were able to explain more than 32% of the variance of the WAIS FSIQ for this sample. The existing formulae had been able to explain much more variance in their original studies. NARTnzBC explained 82% and NARTnzS 42% of FSIQ variance in their original studies. The current FSIQ in these other studies had been measured with a variety of WAIS versions. One of the aims of our study was to compare the scores gained by the different measures of estimating premorbid IQ to the same current IQ measure, here the WAIS-IV. As described in the introduction the WAIS-IV is quite different in many ways to the previous versions. Some of the WAIS-III subtests have been removed (Picture Arrangement), or have become supplementary subtests (Letter-Number-Sequencing and Picture Completion). The subtest of Visual Puzzle appears for the first time in the WAIS-IV. The examination of the changes between the WAIS-III and WAIS-IV, however, did not result in any findings which could explain the lack of accuracy of the existing and new formulae in the prediction of current IQ for the sample in this study.

Examination of the NZART

A second aim of this study was the further validation of the NZART through the use of a different sample. It was found that the NZART was not able to predict the premorbid IQ categories for the scores any more accurately than the other premorbid estimating measures. These results were surprising, because the NZART was developed in New Zealand and the developers had paid attention to all the findings of previous research concerning word familiarity (Barker-Collo, et al., 2008) and cultural suitability (Odgen, et al., 2003), and were careful to include commonly used words from Te Reo Maori to suit Maori participants better.

Having said that the NZART's performance was disappointing it must also be said that the NZART did not perform any worse than the other comparable measures. The explanation for the NZART's performance might lay in the difference between the two samples. The development sample consisted of mainly young student with a mean age of about half of that of the present study (25 compared to 46 years of age). Use of language changes over the years and different age groups are familiar with different words. A mean age gap of 21 years between the two samples could well have a great influence on word familiarity.

This became particularly clear for the Te Reo Maori words. The three Te Reo Maori words used in the NZART were in positions 5 (Maori), 15 (Whenua) and 18 (Kaitiaki) in the NZART word list (Appendix E). Words number 15 and 18 had fewer correct pronunciations than their surrounding words. This indicates that participants in our sample, of which about half where over 45 and white middle class, had more difficulties pronouncing these words correctly than the participants of the developing sample (Starkey & Halliday, 2011). Words, which were a little bit more ‘old fashioned’ such as ‘Lingerie’ (position 17) and ‘Topiary’ (position 53) were pronounced correctly more frequently than their surrounding words, which would indicate that on average our participants found these words easier than their younger counter parts.

The NZART has a few advantages over the NART and TOPF even if its’ performance was very similar to them. The NZART contains at least some New Zealand specific words, it is a little longer than the NART and a little shorter than the TOPF and most importantly, it does not cost anything to use as it has not been commercialised. At this stage and in its’ original order the NZART is possibly better suited for younger participants around the ages of 18 to 50 years old. We recommend a modified order for the NZART based on the findings which would take the word familiarity of older participants into account. With an aging population in New Zealand it is important to provide a measure that reflects the rising age of the test takers.

Limitations

Although the current study attempted to address issues with earlier work, the current study has still a range of limitation. The main issue was the differences between the general population demographics and the sample demographics particularly for age, years in formal education, ethnicity, gender, and location. Further research should endeavour to match the study sample to the population demographics maybe with the use of recent censor data. This would increase the validity of such a study and would clear up some of the issues raised in this study pertaining to age and ethnicity.

Because the age range of the participants was so great, variables like years in education were not necessarily measuring the same construct. The content and quality of education has changed a lot in the last 80 years and access to tertiary institutions has increased dramatically. Only the gathering of an age stratified sample would alleviate these education and age related problems. Sadly, this would be a very costly and time consuming exercise and funding would be very difficult to obtain.

Our study sample had only a small number of Maori participants despite special efforts made to advertise the study through the Maori network at the University. All Maori participants were past or present student of the University which is not a reflection of the general Maori population. To increase the participation of Maori participants researchers from the Maori communities could be conducting similar studies and possibly do Marae based assessment to minimise anxiety due to unfamiliar environments, such as University offices, for the participants.

It was challenging for the researcher to recruit male participants in the older age groups. The community groups contacted by the researcher were mainly comprised of women, and the researcher found it difficult to contact more male orientated groups, possibly because these groups usually meet in the evenings when the researcher could not easily go out to meet them because of family commitments. Future researchers should ensure to advertise a similar study in more male dominated workplaces and groups. It might be valuable to have a male researcher.

The research was conducted in four different locations namely Cambridge, Hamilton, Tauranga and Takapuna in Auckland. All of these locations are quite wealthy and a lot of the participants where well off financially. Because socio economic variables were not collected for this study, the influence of these on the results are not known. Future research would benefit from a sample which resembles the socio economic distribution of the New Zealand population and should endeavour to collect data of the income of the participants in, order to account for the variance due to factors related to socio economics.

A further limitation of our study is inherent to the method of estimating premorbid IQ with the use of regression formulae. This regression towards the mean is a well-known phenomenon and shared by all studies based on regression formulae. Typically, low current IQ scores get over estimated while high current IQ scores get under estimated. There are methods of dealing with this phenomenon but they are very complex and a discussion of these is outside the scope of this thesis. For a review of these please see Veiel and Kooperman (2001) and Graves (2000). Clinicians must be aware of this phenomenon to avoid the wrong estimation of premorbid IQ which could lead to issues with inappropriate funding and treatment of people with brain injuries.

Future Directions

The tests to estimate premorbid IQ used in this study did not predict the current IQ particularly well. All these tests are based on word recognition and the ‘hold/don’t hold’ paradigm which was developed in1954 (Lezak, 2004). These tests are based on the assumption that the chance of exposure to the irregular words in these tests is higher for intelligent people because they have more advanced education in which these words are more likely to be used. The hold/don’t hold paradigm states that some cognitive functions are more resilient to brain damage than others and language related abilities seem to fit into the hold category.

As stated above verbal ability has been used as a premorbid IQ estimator for some time now. It is possible that the assumptions on which these tests are based are now no longer valid. The exposure to the irregular words might not be as stratified as it was 60 years ago as a result of people’s changed reading habits. A well educated person 60 years ago would have spent a considerable amount of time reading books and as a consequence would have maintained their familiarity with these irregular words. Since the appearance of the personal computer a lot of people, educated and otherwise, spend more time reading short, simply worded messages or snippets of information on their computer or mobile phone screens. Irregularly spelt words are no longer very common or in some cases their spelling has been simplified for example ‘night’ to ‘nite’. As a consequence of this the familiarity with irregular words might be decreasing across the population and the ability to pronounce these words is no longer an indicator of extensive education or higher intelligence.

If that is the case then it might be better to use premorbid IQ estimators which are not based on verbal abilities. There is some evidence that the Picture Completion supplementary subtest of the WAIS-IV tests skills which are resistant to brain damage, at least in people who suffered injuries to their left brain hemisphere (Vanderploeg & Schinka, 1995). They used the scores of the 1880 participants of the WAIS-R standardisation sample for each subtest, combined each of these scores with the demographic factors of socio economic status (from occupation and education), ethnicity, gender and age and calculates 33 regression formulae (11 subtests for FSIQ, VIQ and PIQ respectively). They found that Comprehension, Information and Vocabulary subtests were the most stable after right hemisphere damage, while Picture Completion was most stable after left hemisphere damage and recommended the use the formulae most suited to the site of the damage. Others have also found the subtest of Picture Completion resistant to damage (Krull, Scott, & Sherer, 1995; McFie, 1975). Based on these observations and findings future research should examine the abilities of Picture Completion as a predictor of premorbid IQ.

Another approach to estimating premorbid IQ which does not rely on the use of regression formulae and is therefore not in danger of the regression towards the mean phenomenon, is the Best Performance Method (Lezak, 2004). This method relies on the combination of observation, test results, interviews, assessment of past achievements, school records, employment records and so on to find the patient’s best performance level (Lezak, 2004). The assessing clinician decides which information to gather and include and based on these findings will build up a profile of the person before the brain damage which can be compared to the current level of cognitive functioning. The advantage of this method is that the assessment is not restricted to the cognitive abilities tested by the IQ test and can be easily adapted for any individual circumstances. On the other hand the quality of the assessment is very vulnerable to clinician’s subjectivity. This method has often been criticised by researchers because in some studies only the best test score of an IQ test subtest was used as a predictor of premorbid IQ. Of course this is too inaccurate and not what Lezak (2004) proposed.

Because the vocabulary based tests in our study did not perform as well as expected we recommend that future research should look for alternative methods such as the Best Performance Method to explore the used of these for the estimation of IQ in New Zealand.

Unfortunately for the person who experienced a brain injury, there are no clear and easy answers to the questions if they are going to get better and when. However, the neuropsychologist has the above researched tools available to measure the current IQ and estimate the premorbid IQ to gain a valuable insight into the extent of the brain damage and this information coupled with their professional knowledge will provide some of the answers to the patient's questions. The neuropsychologist needs to keep in mind that the estimation of the premorbid IQ will, at least for a while, be just that:- an estimation-; until better, more accurate measures can be developed. It is important, though, to keep in mind that the estimation of the brain damage is only one part of the assessment to answer the questions of the person who suffers the brain damage. Just as IQ tests test only a culturally predetermined part of cognitive function, so the premorbid IQ tests test only a small part of the current cognitive function. There are many more abilities within each person that are not even considered here, which are available to the person with the brain damage to help with the navigation through a world changed as a result of the brain injury.

References

Aron, A., & Aron, E. N. (2006). *Statistics for Psychology* (4th ed.). Upper Saddle River, New Jersey: Pearson Education Inc.

Baddeley, A., Hazel, E., & Nimmo-Smith, I. (1992). *The Speed and Capacity of Language- Processing Test*. London: Pearson Assessment.

Barker-Collo, S., Bartle, H., Clarke, A., van Toledo, A., Vykopal, H., & Willetts, A. (2008). Accuracy of the National Adult Reading Test and Spot the Word Estimates of Premorbid Intelligence in a non-clinical New Zealand sample. [Article]. *New Zealand Journal of Psychology, 37*(3), 53-61.

Barker-Collo, S., Kelly, T., Riddick, E., & de Jager, A. (2011). New Zealand regression formula for premorbid estimation using the National Adult Reading Test. *New Zealand Journal of Psychology, 40*(2), 47- 55.

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Clincal Neuropsychology, 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*(3), 239-250. doi: 10.1016/0887-6177(94)90029-9

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

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic, 34*, 34-37.

Bowden, S. C., Saklfske, D. H., & Weiss, L. G. (2010). Invariance of the measurement model underlying the Wechsler Adult Intelligence Scale-IV in the United States and Canada. *Educational and Psychological Measurement, 71*(1), 186- 199.

Bright, P., Jaldow, E., & Kopleman, M., D. (2002). The National Adult Reading Test as a measure of premorbid intelligence: A comparison with estimates derived from demographic variables. *Journal of the International Neuropsychological Society, 8,* 847- 854.

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since *Bias in Mental Testing*. *School Psychology Quaterly, 14*(3), 208-238.

Calson, J. A. (1995). [Review of the National Adult Reading Test (NART)] *The twelfth mental measurements yearbook*. Available from http://www.unl.edu/buros/.

Canivez, G. L., & Schraw, G. (2010). [Review of Wechsler Adult Intelligence Scale- Fourth Edition] *In the eighteenth mental measurements yearbook*. Available from http://www.unl.edu/buros/.

Coaley, K. (2009). *An introduction to psychological assessment and psychometrics*. London: SAGE Publications Ltd.

Cockburn, J., & Smith, P. T. (2003). *The rivermead behavioural memory test supplement three*. Suffolk, England: The Thames River Valley Test Company Ltd.

Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment: an introduction to tests and measurement* (sixth ed.). New York: Mc Graw Hill.

Crawford, J. R. (1992). *Current and premorbid intelligence measures in neuropsychological assessment*. Hillsdale, New Jersey: Lawrence Earlbaum Associates, Inc.

Crawford, J. R., & Allan, K. M. (1997). Estimating premorbid IQ with demographic variables: Regression equasion derived from a UK sample. *Clinical Neuropsychologist, 11*, 192- 197.

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, M. D., Stewart, L. E., Besson, J. A. O., & De Lacey, G. (1989). Predicion of 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. (1989a). Construct validity of the National Adult Reading Test: a factor analytic study. *Personality and Individual Differences, 10*(5), 585-587. doi: 10.1016/0191-8869(89)90043-3

Crawford, J. R., Stewart, L. E., Cochrane, R. H. B., Parker, D. M., & Besson, J. A. O. (1989b). Cunstruct validity of the National Adult Reading Test: a factor analytic study. *Personaliity and Individual Differences, 10*(5), 585- 587.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2009). *Advanced Clinical Solution for WAIS-IV and WMS-IV Administation and Scoring manual*. San Antonio, TX: The Psychological Corporation.

Field, A. (2009). *Discovering statistics using SPSS* (Third edition ed.). London: SAGE Publications Ltd.

Flynn, J., R. (2009). The WAIS-III and WAIS-IV: Daubart Motion favors the certainly false over the aproximately true. *Applied Neuropsychology, 16*, 98- 104.

Franzen, M., D, Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology, 12*(8), 711- 738.

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. [Article]. *British Journal of Clinical Psychology, 40*(2), 221.

Gladsjo, J. A., Heaton, R. K., Palmer, B. W., Taylor, M. J., & Jeset, D. V. (1999). Use of oral reading to estimate premorbid intellectual and neuropsychological functioning. *Journal of the International Neuropsychological Society, 5*(03), 247-254. doi: doi:null

Graves, R. E. (2000). Accuracy of regression equation prediction across the range of estimated premorbid IQ. *Journal of Clincal and Experimental Neuropsychology, 22*(3), 316- 324.

Green, R., E, Melo, B., Christensen, B., Ngo, L.-A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology, 30*(2), 163- 172.

Groth- Marnat, G. (2009). *Handbook of Psychological Assessment*. Hoboken, New Jersey: John Wiley& Sons,Inc.

Hartman, D. E. (2009). Test review Wechsler Adult Intelligence Scale-IV (WAIS-IV): Return of the gold standard. *Applied Neuropsychology, 16*, 85- 87.

Homer. (2003). The Odyssey. New York: Barnes & Nobel Classics.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340-362. doi: 10.1016/0001-8791(86)90013-8

Hutchins, R. M. (Ed.). (1952a). *Great Books of the Westen World: Kant, Critique of Pure Reason* (Vol. 42). Chicago: Encyclopeadia Britannica.

Hutchins, R. M. (Ed.). (1952b). *Great Books of the Western World : Machiavelli, Hobbes. The Prince by Nicolo Machiavelli; Leviathan by Thomas Hobbes* (Vol. 23). Chicago: Enceclopaedia Britannica.

Hutchins, R. M. (Ed.). (1952c). *Great Books of the Western World: Aquinas, Summa Theologica* (Vol. 19). Chicago: Enceclopeadia Britannica.

Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist, 9*, 83- 88.

Lezak, M., D. (2004). *Neuropsychological assessment*. Oxford: Oxford University Press.

Lichtenberger, E. O., & Kaufman, A. S. (2009). *Essentials of WAIS-IV Assessment*. Hoboken, New Jersey: John Wiley& Sohn Inc.

Loring, D. W., & Bauer, R. M. (2010). Testing the limits: Cautions and concerns regarding the new Wechsler IQ and Memory scales. *Neurology, 74,* 685-690.

Mathias, J. L., Bowden, S. C., & Barrett-Woodbridge, M. (2007). Accuracy of the Wechsler Test of Adult Reading (WTAR) and National Adult Reading Test (NART) when estimating IQ in a healthy Australian sample. [Article]. *Australian Psychologist, 42*(1), 49-56. doi: 10.1080/00050060600827599

McFie, J. (1975). *Assesment of organic inellectual inpairment*. New York: Academic Press.

McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., . . . Deary, I. J. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology, 62*, 1184-1186.

Neisser, U., Boodoo, G., Bouchard, T., Brody, N., Ceci, S. J., Halpern, D. F., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77- 101.

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART)* (2nd ed.). Berkshire, Great Britain: NFER- Nelson Publishing Company Ltd.

Odgen, J., A., Cooper, E., & Dudley, M. (2003). Adapting neuropsychological assessments for minority groups: A study comparing white and Maori New Zealanders. *Brain Impairment, 4*(2), 122- 134.

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

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom:Teacher expectations and pupils' itellectual development.* New York: Holt, Rinehart and Winston.

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

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

Smyth, T. R. (2004). *The principles of writing in psychology*. New York: Palgrave MacMillan.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests*. New York: Oxford University Press.

Starkey, N., J, & Halliday, T. (2011). Development of the Newe Zealand Adult Reading Test (NZART): Preliminary findings. *New Zealand Journal of Psychology, 40*(3), 129- 141.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. Cambridge: Cambridge Universirty Press.

Stobart, G. (2008). *Testing times: The use and abuse of assessment*. London: Routledge.

Thompson, N. M., & Ward, S. (2001). [review of the Wechsler Test of Adult Reading] *The sixteenth mental measurements yearbook.* Available from http://www.unl.edu/buros/.

Uttl, B. (2002). North American Adult Reading Test: Age Norms, Reliability, and Validity. [Article]. *Journal of Clinical & Experimental Neuropsychology, 24*(8), 1123.

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

Vanderploeg, R. D., & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid ability: Combining subtest performance and demographic variable predictors. *Archives of Clinical Neuropsychology, 10*(3), 225-239. doi: 10.1016/0887-6177(94)00042-o

Veiel, H. O. F., & Kooperman, R. F. (2001). The bias in regression- based indices of premorbid IQ. *Psychological Assessment, 13*(3), 356- 368.

Watkins, J., & Hogenhout, E. (2010). *National population estimates: March 2010 quater*. Christchurch, New Zealand.

Watt, K. J., & O'Carroll, R. (1999). Evaluating methods for estimating premorbid intellectual abilities in closed head injury. *Journal of Neurology, Neurosurgery and Psychiatry, 66*(4), 474- 779.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.

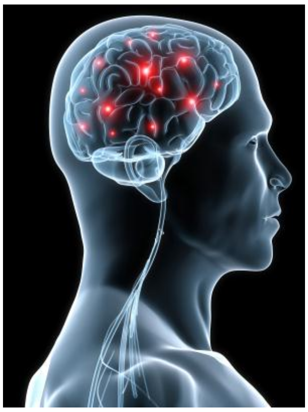
Wechsler, D. (2008). *Wechsler Adult Intelligence Scale- Fourth Edition: Technical and interpretative manual*. San Antonio: The Psychological Corporation.

Wilson, B., A. (1999). *Case studies in neuropsychological rehabilitation*. Oxford: Oxford University Press.

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

Appendix A

The poster used to advertise for participants for the study. (Overleaf)



Interested in Neuropsychology?

Would you be willing to take part in a study to develop neuropsychological tests more appropriate for New Zealand?

If so, you might be interested in taking part in a study being conducted by Irene Lichtwark from the School of Psychology at University of Waikato.

This study, titled "Accuracy of Premorbid Estimations in New Zealand" is recruiting individuals who speak English as a first language, and were born in New Zealand.

Participants will participate in 1 or 2 assessment session of up to 210 minutes at the University of Waikato or in your home. The assessment will involve administration of test commonly used by psychologists in New Zealand. There are no risks associated with this study, many people find the tasks interesting, and it is likely that you may find some of the items difficult.

Participation is **voluntary;** and all information collected will remain **confidential**. That is, your name/identifying information will not be associated with published results.

A $20 voucher (or course credit for first year psychology students) will be given to participants as a token of appreciation of their participation, regardless of whether they decide to withdraw. The study has received approval from the School of Psychology Ethics Committee and is supervised by Dr Nicola Starkey Tel: 07 8562889 or email: nstarkey@waikato.ac.nz .

If you wish to participate or would like further information please contact me. Irene Lichtwark, e mail: itl4@waikato.ac.nz or phone 0212347740.
I am looking forward to hearing from you.

Appendix B

The demographic forms developed for this study.

Demographic Information Sheet

General Information

Age: ________________________ Gender: ____________________

Years of Education: ___________ Marital Status: _______________

Years in New Zealand: __________

Occupation (eg student, lawyer): See other form ____________________

Ethnicity

How would you describe your culture of origin: (e.g., European, Tongan, Maori):

____

Cultural Identity:

Which of the following cultures do you *most* identify yourself with?
(please check only one box)

Maori ☐ Pakeha ☐ Other : ________________(specify)

Other Factors

Please indicate if you have ever received any of the following diagnoses or classes of diagnoses:

Concussion	Yes	No
Head Injury	Yes	No
Learning Disability	Yes	No
Depression	Yes	No
Anxiety Disorder	Yes	No
Personality Disorder	Yes	No
Psychotic Disorder	Yes	No

Are there any other factors that may impact on your performance/responses? (Alcohol in last 24 hours, etc)

____

Is English your first language? Yes No

If No, what is your first language (specify): ______________________

Handedness ________________________

Appendix C

Figures showing the sums of correct pronunciations for the TOPF and NZART.

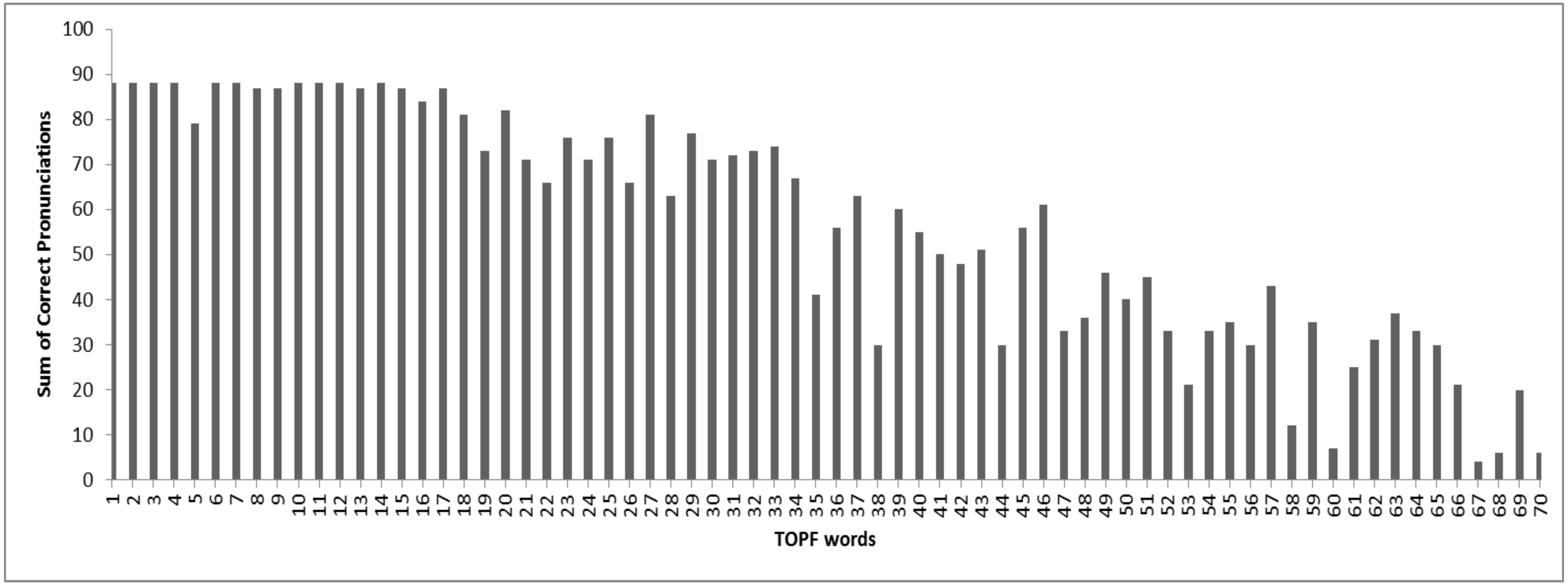


Figure C1. The sum of correct pronunciations for each word of the TOPF. The words are in their original order.

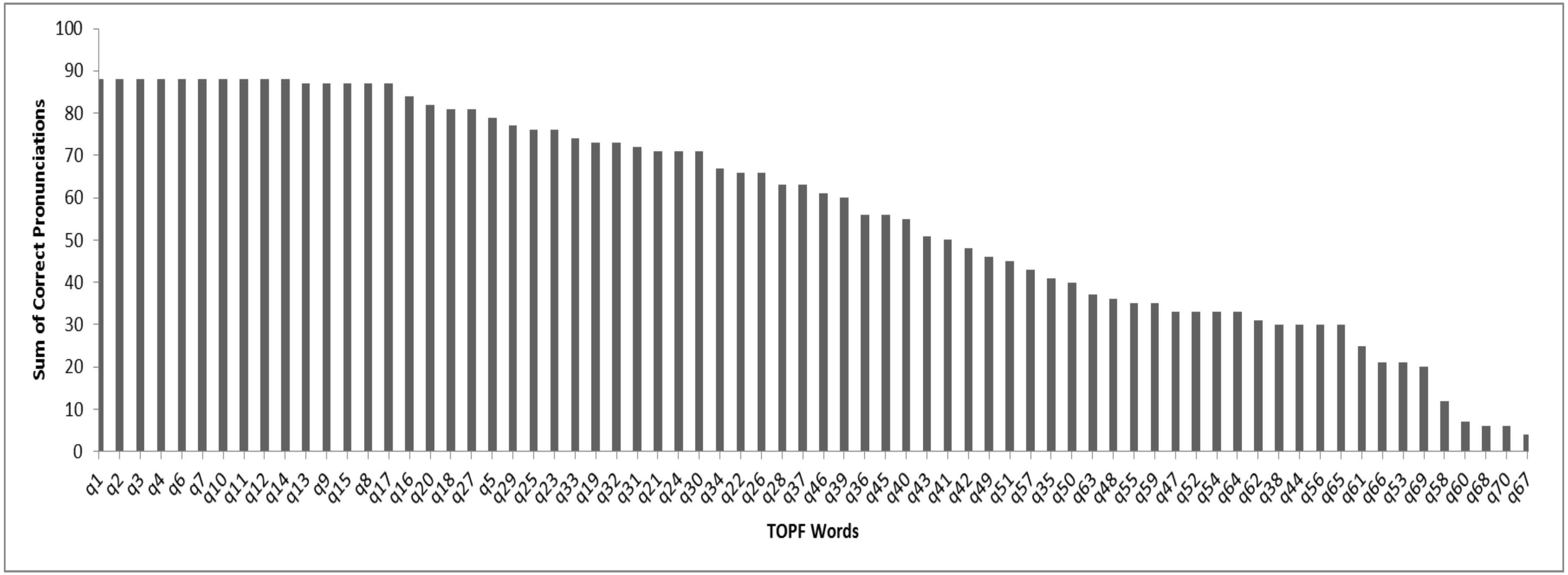


Figure C2. The sum of correct pronunciations for each word of the TOPF. The words are in their new order based on the sum of correct scores.

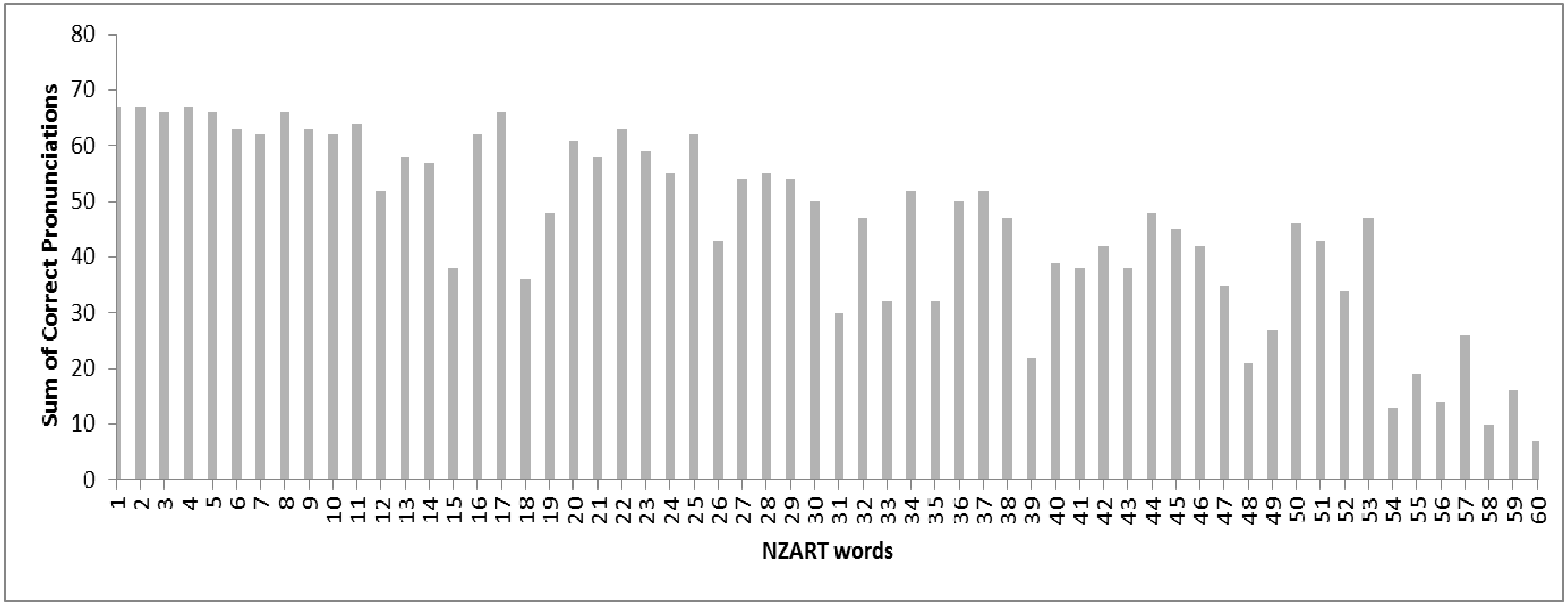


Figure C3. The sum of correct pronunciations for each word of the NZART. The words are in their original order.

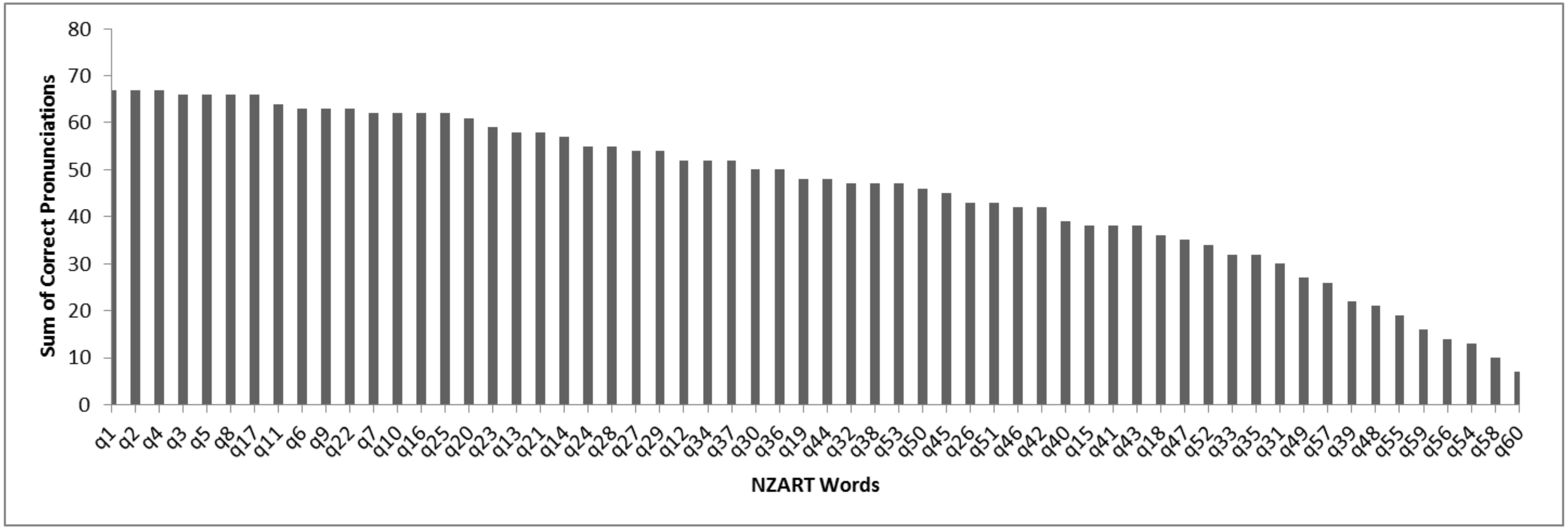


Figure C4. The sum of correct pronunciations for each word of the NZART. The words are in their new order based on the sum of scores.

Appendix D

Tables showing the development of the regression formulae for WAIS VCI using the NART (Table D.1), the TOPF (Table D.2.) and the NZART (Table D.3.) error scores.

The regression calculated with WAIS-VCI as the dependent variable and NART and education as the predictors. Again the entry was forced to explore if education was able to improve the prediction based on NART error scores alone. Table D.1 shows the results.

Table D.1

Results of the Forced Entry Regression with WAIS VCI as the Dependent Variable and NART Error Scores and Education as Predictors.

	B	*SE B*	*beta*
Step 1			
Constant	129.67	3.49	
NART errors	-1.00	.16	-.57***
Step 2			
Constant	111.83	6.50	
NART error	-.91	.16	-.52***
Education in years	1.10	.34	.29**

Note. R^2 = .32*** for Step 1, delta R^2=.08** for Step 2,
**p<.01, ***p<.001

The NART alone predicted a significant amount of variance in VCI (R^2=.32; F(1,78) =37.08, p<.001), accounting for 32.2% of the variance in VCI. The addition of the demographic factor significantly improved the model (R^2_{change}=0.8; F_{change}(1,77) =10.18, p<.01), accounting for an overall variance in VCI of 40.1 % with a standard error of estimate of 9.51. Contributing significantly to the prediction were the NART error score (p<.001) and years of education (p<.01). The resulting formula for the prediction of VCI was:

NART Predicted VCI= 111.83+ (-.91 NART error score) + (1.10 education)

The next stepwise regression calculated was for WAIS VCI as the dependent variable and TOPF scores and education in years as predictors. Table D.2 reports the results.

Table D.2

Results of the Forced Entry with WAIS VCI as the Dependent Variable and TOPF Scores and Education as Predictors.

	B	*SE B*	*beta*
Step 1			
Constant	87.91	4.65	
TOPF score	.46	.10	.48***
Step 2			
Constant	73.24	6.34	
TOPF score	.41	.09	.43***
Education in years	1.20	.36	.31**

Note. R^2 = .23*** for Step 1, delta R^2=.09** for Step 2, **p<.01, ***p<.001

The TOPF alone predicted a significant amount of variance in VCI (R^2=.23; F(1,78) =23.18, p<.001), accounting for 22.9% of the variance in VCI. The addition of the demographic factor significantly improved the model (R^2_{change}=.09; F_{change}(1,77) =10.34, p<.01), accounting for an overall variance in VCI of 32.0% with a standard error of estimate of 10.13. Contributing significantly to the prediction were the TOPF score (p<.001) and years of education (p<.01). The resulting formula for the prediction of VCI was:

TPOF Predicted VCI= 73.24+ (.41 TOPF score) + (1.20 education in years)

The last regression computed was again a stepwise entry regression to show why the predictor variable of education in years was excluded from the final formula and Table D.3 shows the results.

Table D.3

Results of the Stepwise Multiple Regression with WAIS VCI as the Dependent Variable and NZART Error Scores and Education as Predictors.

	B	*SE B*	*beta*
Step 1			
Constant	122.55	2.67	
NZART errors	-.69	.13	-.57***
Step 2			
Constant	109.47	6.70	
NZART error	-.64	.12	-.52***
Education in years	.82	.39	.22*

Note. R^2 = .32*** for Step 1, delta R^2=.04* for Step 2,
**p<.01, ***p<.001

The NZART alone predicted a significant amount of variance in VCI (R^2=.32; F(1,65) =30.92, p<.001), accounting for 32.2% of the variance in VCI. The addition of the demographic factor did significantly improved the model (R^2_{change}=.04; F_{change}(1,64) =4.49, p>.05), increasing the overall variance explained for VCI to 36.7% with a standard error of estimate of 10.05. Contributing significantly to the prediction were the NZART error score (p<.001) and education in years (p<.05). The resulting formula for the prediction of VCI was:

NZART Predicted VCI= 109.47+ (-.64 NZART errors) + (.82 education in years)

Appendix E

The word list of the NZART. (overleaf)

NZART

DEBT	**WHENUA**	**TACIT**
CHOIR	**THYME**	**COLONEL**
AISLE	**LINGERIE**	**REIFY**
CHAOS	**KAITIAKI**	**COGNAC**
MAORI	**INSATIABLE**	**AMYGDALOID**
NAUSEA	**COURTEOUS**	**RISQUE**
GROTESQUE	**HIATUS**	**EPITOME**
FATIQUE	**MERINGUE**	**INDICES**
COLOGNE	**DEBRIS**	**CHASSIS**
SUBTLE	**INERTIA**	**SUPERFLUOUS**
NAIVE	**PLACEBO**	**LEVIATHAN**
PSALM	**CHAMELEON**	**SUBPOENA**
TORQUE	**EQUIVOCAL**	**FACETIOUS**
SIEVE	**CROCHET**	**OCHRE**

IMPUGN	**INDICT**	**VIVACE**
ZEALOT	**CAVEAT**	**LABILE**
FACADE	**CORPS**	**DETENTE**
TOURNIQUET	**ABSTEMIOUS**	**CAECUM**
HIPPOCRATES	**TOPIARY**	**TALIPES**
QUADRUPED	**IDYLL**	**SYNCOPE**