Accuracy of the National Adult Reading Test and Spot the Word Estimates of Premorbid Intelligence in a non-clinical New Zealand sample

Suzanne Barker-Collo, Hamish Bartle, Annabel Clarke, Annik van Toledo, Helen Vykopal, Amanda Willetts, *Department of Psychology, The University of Auckland*

This study investigated National Adult Reading Test (NART) and Spot the Word (STW) premorbid estimates compared to Wechsler Adult Intelligence Scale (3rd ed; WAIS-III) performances in 89 healthy New Zealand adults. STW and NART Full Scale Intelligence (FSIQ) estimates were strongly correlated to WAIS-III FSIQ for Europeans (n=75; rNART = .70, *p* < .001; *r* STW = .70 *p* < .001). For those who self-identified as being of Māori ancestry (n=14) there was no significant correlation between WAIS-III FSIQ and the NART, while the correlation between STW and WAIS-III FSIQ was large (*r* = .91, *p* < .001). For Europeans the NART accurately predicted 41% of classifications, compared to 52% for STW. For Māori STW correctly classified 93%, compared to 7% for the NART. Though study with larger samples is required, when used in conjunction with qualitative information about ability, STW may be a more accurate premorbid estimator for New Zealand.

Estimates of premorbid intelligence are an essential aspect of neuropsychological assessment, with the contrast between premorbid level of functioning and current level of performance forming the basis of judgements about areas of preserved versus impaired functioning (Baddeley, Emslie, & Nimmo-Smith, 1993; Klesges & Troster, 1987; Lezak, Howieson, & Loring, 2004; Vanderploeg & Schinka, 1995; Watt & O'Carroll, 1999). As people are not often assessed prior to disease/injury onset, no direct comparison to premorbid functioning can be made, and indirect estimations of premorbid abilities are required (Bright, Jaldow & Kopelman., 2002; Hart, Smith, & Swash, 1986; Klesges & Troster, 1987; O'Carroll, 1995; Petito, 2000). Predictive methods must provide estimations that correlate highly with actual functioning (typically measured as intelligence [IQ]) in normal populations and must also be resilient to decline in cognitive function (Schoenberg, Scott, Duff, & Adams, 2002; Sharpe & O'Carroll, 1991). Such methods include demographic based regression equations, and use of tests of overlearned skills thought to be resilient to brain damage (Lezak, et al., 2004). In New Zealand, regression formulae to estimate premorbid intelligence are, as yet, unavailable. In clinical practice, judgements about premorbid level of functioning are often made using qualitative information about overall ability in conjunction with tests of overlearned skills such as the National Adult Reading Test (NART) and Spot the Word (STW). As such, the literature around these will each be reviewed briefly below.

The NART and STW

The NART (Nelson & Willison, 1991) is one of the most commonly used premorbid estimation methods. It consists of 50 irregularly spelled words that examinees are asked to pronounce correctly, as a measure of reading ability (Collins, 2000; Crawford, Parker, Stewart, Besson, & De Lacey, 1989; Hart et al., 1986; Nelson & McKenna, 1975; Sharpe & O'Carroll, 1991). Reading ability is a potentially useful method of predicting overall functioning as it requires a complex interaction of cognitive functioning, with inclusion of irregular words ensuring that it tests familiarity of a particular word rather than the ability to work out words phonetically (Collins, 2000; Nelson & O'Connell, 1978). A regression equation to predict premorbid IQ from the NART has been developed (Nelson & McKenna, 1975), and is a better predictor of premorbid functioning than the WAIS-R Vocabulary subtest (Collins, 2000; Petito, 2000). In a retrospective study, NART IQ estimates in people aged 77 years correlated significantly with IQ scores obtained from them at age 11 years, indicating only minimal effects of age and life experience on NART estimates (Crawford, Deary, Starr, & Whalley, 2001). The NART also remains highly stable over time (Morrison, Sharkey, Allardyce, Kelly & McCreadie, 2000; Smith, Roberts, Brewer, & Pantelis, 1998), and has been reported to be an effective predictor of premorbid IQ in closed head injury and orthopaedic trauma patients (Watt & O'Carroll, 1999). However, some authors have shown the NART to be sensitive to changes in cognitive functioning such as in patient's with Alzheimer's disease (Cockburn, Keene, Hope & Smith, 2000) or in the acute phase after severe traumatic brain injury (TBI) (Riley & Simmonds, 2003). Further, the NART cannot be used in patients suffering from dyslexia, dysarthria, visual or articulatory problems, and it underestimates IQ in self-educated populations (Baddeley et al., 1993; Crawford, Cochrane, Besson, Parker, & Stewart, 1990b) and people with depression (Watt & O'Carroll, 1999).

The STW (Baddeley, Hazel, & Nimmo Smith, 1992) was developed as part of the Speed and Capacity of Language Processing (SCOLP) assessment (Saxton, Ratcliff, Dodge, Pandau, Baddeley, & Ganguli, 2001) and consists of 60 word pairs. Examinees must decide which word in each pair is a real word, and which a non-word. This format that allows lexical decisions to be made through multiple methods including meaning, familiarity, appearance and sound of words (Baddeley et al., 1993; Saxton et al., 2001). Studies have shown that performance on the STW is not impacted by age or gender, though increased education does result in higher scores (Baddeley et al., 1993; Saxton et al., 2001; Yuspeh & Vanderploeg, 2000). It is also relatively stable following neurological impairment and is a good estimator of premorbid abilities as it shows good discriminant validity (Yuspeh & Vanderploeg, 2000). As no verbal response is needed, it also has the advantage that it can be used in patients who suffer from dyslexia, expressive aphasia or articulatory problems (Baddeley et al., 1993).

Cross Cultural Validation

As both the NART and STW were originally developed in Britain, if used outside Britain they are potentially sensitive to cultural effects due to differences in lexicon, pronunciation, and word frequency differences between populations (Franzen, Burgess, & Smith-Seemiller, 1997; Paolo et al., 1997). Cultural differences can have important consequences for neuropsychological assessment, whether this is a result of differences in anxiety levels, test taking strategies or familiarity of content between cultural groups (Brown, Reynolds, & Whitaker, 1999; Franzen et al., 1997; Freeman, Godfrey, Harris, & Partridge, 2001; Helms, 1992; Knight, McMahon, Green, & Skeaff, 2004; Kohnert, Hernandez, & Bates, 1998). Thus, if comparisons are made to inappropriate standardisation samples, premorbid estimations are likely to be inaccurate, leading to incorrect diagnoses, and thereby recommendations for rehabilitation.

In New Zealand these factors may be particularly salient to performance on verbal tests, which has implications for the use of tests such as the NART and STW. A number of previous studies have documented that New Zealanders in general (Barker-Collo, 2001; Barker-Collo, Clarkson, Cribb, & Grogan, 2002) and Māori more specifically (e.g., Ogden & McFarlane-Nathan, 1997) perform worse on tests requiring verbal abilities than would be expected when compared to existing (overseas) normative data. Two of these studies (Ogden & McFarlane-Nathan, 1997; Barker-Collo, et al., 2002) have produced evidence that removing unfamiliar words from these tests and substituting words that are more relevant to a New Zealand context can improve performance to a level consistent with standardisation samples.

In regards to the NART and STW in New Zealand, Freeman et al. (2001) conducted a study of the NART in New Zealand samples, finding that almost a third of a sample of TBI patients had lower NART performance in comparison to expected performance based on demographic equations developed by Crawford, Allan, Cochrane, and Parker (1990a). Franzen et al. (1997) have suggested that such differences are likely due to differences in word familiarity, indicating that use of the NART outside Britain may not be appropriate without some modification. Thus, validating the NART in a healthy New Zealand sample is important for its continued utility in clinical settings. In an as yet unpublished study of 32 healthy New Zealand European and 21 healthy Māori University students, Halliday (2006) reports significant correlations between NART and Wechsler Abbreviated Intelligence Scale (WASI; Wechsler, 1999) estimated IQ scores. In the only New Zealand study found to include the STW, it was found that 78% of Māori and 56% of Europeans did not like the STW measure, even though in the past it has been suggested that STW might be less stressful for patients than the NART (Ogden, Cooper & Dudley, 2003). Unfortunately this study does not report on the relationship between STW and overall IQ in this sample.

The above highlights the importance of cultural considerations in neuropsychological assessment, including premorbid estimates, in New Zealand. The purpose of this study was to examine the relationships between WAIS-III IQ scores and both NART and STW IQ estimates in a non-clinical, New Zealand born sample.

Method

Participants

The 89 New Zealand born participants, 45 (50.6%) of whom were male and 44 (49.4%) female, ranged in age from 19 to 67 years (mean = 33.64; *sd* = 12.87). Years of education in the sample ranged from 9 (intermediate school) to 26 years (postgraduate degree; mean = 15.10; *sd*= 2.80). Time spent living outside of New Zealand ranged from 0 to 15 years (mean = 1.36; *sd*= 3.36). The majority of participants were single (*n* = 61; 67.8%), with 17 (18.9%) of the remainder being married, 7 (7.8%) in de facto relationships, and 4 (4.4%) separated or divorced. Seventy-five participants (83.3%) were of European ancestry, with the remaining 14 (15.6%) self-identifying as Māori. Most (*n* = 48; 53.3%) of the participants were 100% right handed on the Edinburough Handedness Inventory, while 37 (41.1%) were 100% left-handed, and 4 (4.4%) were ambidextrous. Individuals were excluded from the sample if they experienced a current psychiatric disorder, did not speak English as a first language, or had a history of head injury or other neurological condition. Five (5.6%) participants reported a history of possible concussion not resulting in loss of consciousness or medical treatment. In regards to employment, 12 (13.5%) of participants were in unskilled work (e.g., clerical, labourer), 22 (24.7%) were skilled workers (e.g., chef, plumber), 26 (29.2%) were professionals (e.g., biologist, administrator), and the remaining 29 participants (32.6%) were tertiary students.

Measures

National Adult Reading Test (NART; Nelson & Willison, 1991). The NART consists of a list of 50 unrelated, phonetically irregular words of graded difficulty. While an American version of the task is available, New Zealand's closer ties to Britain mean that this is the version most commonly used in New Zealand, and was therefore used in this study. Participants are required to read aloud the word list in sequential order and are scored on correct pronunciation with a maximum score of 50. Crawford et al. (2001) demonstrated good criterion validity with the NART, explaining between 55% and 72% of the variance in IQ. Reliability is shown by good internal consistency (α= 0.90; Crawford, Stewart, Garthwaite, Parker, & Besson, 1988), inter-rater reliability *(r* = 0.96 and 0.98) and test-retest reliability *(r* = 0.98; Crawford, et al., 1989). NART raw scores were converted into IQ estimates in accordance with the NART manual.

Spot the Word (STW; Baddeley, Emslie & Nimmo-Smith, 1992). The STW portion of the SCOLP consists of 60 items. Each item presents the participant with one real word and one pseudo-word (e.g., kitchen - harrick), and he/she must identify the real word from each pair. A total score that reflects the total number of items correct, and scores can range from 0 to 60. The STW is thought to provide a promising estimate of performance intelligence (Baddeley, et al., 1992). STW total scores were converted into standard scores using age-corrected normative tables presented in the manual. As these standard scores have similar psychometric characteristics to standard scores provided by the WAIS-III (i.e. mean = 10, *SD* = 3), it was then possible to convert standard scores to IQ score estimates. STW has adequate reliability (0.83), and validity as assessed by correlating performance on STW with the Mill Hill Vocabulary Scale *(r* = 0.72; Baddeley, et al., 1992).

Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997). The WAIS-III is a battery of neuropsychological subtests assessing various aspects of cognition. The battery contains 14 subtests, 11 of which provide Full Scale IQ (FSIQ), Verbal IQ (VIQ) and Performance IQ (PIQ) scores. These 11 subtests were administered and scored in accordance with standardised procedures (Weschler, 1997). The WAIS-III has a reliability coefficient of 0.98 (Lezak et al., 2004). Correlation between the WAIS-III and the Stanford-Binet Intelligence Scale – Fourth Edition (SB-IV; Thorndike, Hagen & Sattler, 1986) was established as 0.88 (Thorndike et al., 1986) suggesting good validity.

Procedure

Potential participants were identified via student networks and through community agencies that had given approval for the research to be presented to its members (e.g., Age Concern). Potential participants were initially given a general description of the research project verbally. In community agencies this was provided in a small group format. Those interested in participating were then provided with a participant information sheet (PIS) and a consent form that included space in which to provide contact details, and a freepost envelope addressed to the researcher in which to return completed consent forms. Of the 105 potential participants who made contact with the researcher, nine did not meet inclusion criteria, and seven did not consent to participate in the study due to concern about the time commitment required, resulting in a recruitment rate of 85%.

Individuals who agreed to participate were then contacted by the researcher to check eligibility and schedule a testing session at their convenience. Sessions took place either in the participant's home or in an office space on university premises (at participant request) and took approximately 180 minutes to complete, including a 10 minute break. Sessions began with a review of the PIS including the aims of the study, the voluntary and confidential nature of the study, the storage and use of test data, and the expected time commitment. All participants were offered the opportunity to ask questions about the study. Distractions were minimised in all testing sessions by ensuring a quiet environment without interruption (e.g., switching off telephones). All participants completed the NART, STW, and the 11 subtests of the WAIS-III. Order of presentation was counterbalanced with half the participants completed the WAIS-III first and half completed the WAIS-III last. Performance of each participant was scored in accordance with standardised procedures and all data was then entered into a Statistical Package for Social Sciences 15.0 (SPSS; 2007) file for analysis.

Results

The results of this study are presented in three sections. First, overall performance of the sample across the measures (i.e., means and standard deviations) is presented. This is followed by an examination of the impact of demographic grouping variables (i.e., gender and ethnicity) and order of testing (i.e., WAIS-III first versus last) on performance. The remainder presents information on the relationships between WAIS-III performance and performance on the STW and NART. These relationships are first examined as correlations between primary scores on these measures. To better gauge the accuracy of predictions made using STW and NART, the number and percentage of New Zealand Europeans and Māori who's WAIS-III FSIQ fell into each qualitative category and corresponding allocations of NART and STW qualitative categories are examined.

Overall Performance

Means and standard deviations across scores from the WAIS-III, STW and NART are presented in Table 1. As can be seen from Table 1, WAIS-III subtest scores for this sample were within 1 *SD* above the normative mean for Picture Completion, Similarities, Block Design, Digit Span, Information, and Picture Arrangement; between 1 and 2 SDs above the normative mean for Vocabulary, Matrix Reasoning, and Comprehension; and on the mean for Digit Symbol Coding. These scores suggest a sample that was above average, with particular strength in the area of Vocabulary; which was reflected in NART and STW performance. Overall WAIS-III performance of this sample fell within the high average range, as did NART and STW estimates of overall performance. It is notable that while WAIS-III IQ scores and STW IQ estimates had SDs in line with those in the normative population (i.e., 15), this variability was reduced to less than half for the NART estimates.

Table 1.
Means and standard deviations across scores.

Scaled Scores/IQ	Mean[a]	Standard Deviation
WAIS Subtest Standard Scores		
Verbal Subscales		
Vocabulary	13.87	2.81
Similarities	12.17	2.85
Arithmetic	10.82	2.92
Digit Span	11.04	2.81
Information	12.17	3.12
Comprehension	13.26	2.77
Performance Subscales		
Picture Completion	11.49	2.78
Digit Symbol Coding	10.02	3.81
Block Design	12.97	3.15
Matrix Reasoning	13.26	2.28
Picture Arrangement	11.60	3.53
Verbal IQ	113.98	15.17
Performance IQ	113.77	15.92
Full Scale IQ	114.84	15.25
NART		
Raw Total Errors	17.27	6.64
Verbal IQ	113.13	6.10
Performance IQ	112.36	4.28
Full Scale IQ	113.44	5.48
STW		
Raw Total Score	51.77	4.23
Standard Score	12.33	2.66
IQ estimate	111.46	13.06

[a] Standard scores have mean = 10 and *SD* = 3; IQ scores have mean = 100 and *SD* = 15.

Impact of Grouping Variables

Before conducting primary analyses the extent to which performance was impacted by order of test-taking, gender, and ethnicity within the sample were examined. A 2 x 2 x 2 MANOVA was performed with test order (WAIS first; WAIS last), gender, and ethnicity (European, Māori) as grouping variables, and WAIS-III, NART and STW IQ scores as dependent variables. The results of this analysis indicate that there were significant main effects for gender ($F(7, 75) = 3.223$, $p < .001$) and for ethnicity ($F(7, 75) = 2.093$, $p = .019$). No other main effects or interactions were significant ($p > .05$).

Contributing to the significant main effect of gender was WAIS-III VIQ ($p = .029$). Mean scores for males and females were 118.78 (*SD* = 16.13) and 109.07 (*SD* = 12.29), respectively, indicating that males in this sample performed better across WAIS-III verbal tests. A further one-way ANOVA was run to determine which verbal subtest might be contributing to these differences. It was found that males produce significantly higher scaled scores than females on Similarities ($F(2, 80) = 8.717$, $p = .004$; means ♀= 11.59, ♂= 12.73), Arithmetic ($F(2, 80) = 5.099$, $p = .027$; means ♀= 9.57, ♂= 12.04), and Information ($F(2, 80) = 8.608$, $p = .004$; means ♀= 10.95, ♂= 13.36).

All three WAIS-III IQ scores contributed to the differences found between ethnic groups ($F(7, 75)$FSIQ = 9.355, $p = .003$; $F(7, 75)$VIQ = 8.314, $p = .005$; $F(7, 75)$PIQ = 6.620, $p = .012$), as did the STW IQ estimate ($F(7, 75) = 8.582$, $p = .004$). When mean scores were examined, it was found that Europeans produced higher mean scores across VIQ, PIQ FSIQ , and STW (means = 115.96, 116.36, 117.2, and 113.67 respectively) than Māori (means = 103.36, 100.00, 102.21, and 99.64, respectively). In examining between group differences on WAIS-III subtest scaled scores, significant differences were found between European and Māori on Vocabulary ($F = 10.124$, $p = .002$), Similarities ($F = 9.824$, $p = .002$), Comprehension ($F = 8.097$, $p = .006$), Matrix Reasoning ($F = 12.92$, $p = .001$), and Picture Arrangement ($F = 4.575$, $p = .035$). The 14 Māori in the sample obtained lower mean scores across these measures than Europeans (i.e., Vocabulary means = 11.93 and 14.23;, Similarities means = 10.14 and 12.55, Comprehension means = 11.43 and 13.60, Matrix Reasoning means = 10.79 and 13.72, and Picture Arrangement means = 9.28 and 12.03, respectively). One way ANOVA revealed that the two ethnic groups did not differ significantly in terms of age or educational attainment ($p > .05$), nor did Mann Whitney U test reveal significant differences for gender, history of possible concussion, or job category (i.e., student, unskilled labour, skilled, professional) ($p > .05$). Despite the small size of the Māori group within the sample, while these findings need to be viewed with some caution, the number and significance of these

Table 2.

Correlations between predicted and actual IQ scores and age and education for European and Māori groups.

IQ score/Estimate	New Zealand European (N=75) Age	Education[a]	Māori (N=14) Age	Education
WAIS-III				
FSIQ	-.31**	.61**	ns	ns
VIQ	-.43**	.56**	-.78**	ns
PIQ	ns	.44**	ns	ns
NART				
Total score	ns	.50**	ns	ns
IQ[b]	ns	.51**	ns	ns
STW				
Total Score	ns	.38**	ns	ns
IQ	-.23*	.42**	ns	ns

[a]Education was assessed as number of years of education successfully completed

[b]As all NART IQ estimates produce the same correlations these are only provided once.

* $p <.05$

** $p < .01$

ns $p > .05$

differences suggested that ethnicity be considered in all further analyses.

Given the above differences it was also decided to examine, through bivariate correlations, whether the expected relationships between age, education levels and performance held true within this sample. The results of these analyses are presented separately for Māori and European groups in Table 2. As can be seen in Table 2, while the expected correlations were found linking more years of education to better test performance for European New Zealanders, the same was not true for Māori. For Europeans, increased age was also related to worse WAIS-III and STW FSIQ, and WAIS-III VIQ. The only significant correlation found within the Māori sample indicated that younger people performed better on the WAIS-II VIQ. None of the other correlations generated were significant ($p > .05$).

WAIS-III, NART, and STW relationships

To begin exploring the direction and degree of relationships between the NART and STW IQ estimates and WAIS-III IQ scores, bivariate correlations were generated separately for Māori and European groups and are presented in Table 3. As can be seen from Table 3, for New Zealand Europeans the NART total score and STW IQ estimate had the largest correlations with obtained WAIS-III FSIQ, with each explaining 49% of the variance in WAIS-III FSIQ. The STW IQ estimate also had the largest correlation with obtained WAIS-III VIQ, explaining 56% of the variance in VIQ; though it had the weakest relationship to PIQ, explaining only 22% of the variance in obtained WAIS-III PIQ. Correlations for Māori in the sample were in stark contrast, with only the STW FSIQ estimate being significantly correlated to obtained WAIS-III FSIQ, VIQ, and PIQ scores; accounting for 83%, 56%, and 34% of the variance in WAIS-III IQs, respectively.

While the above suggests that NART total scores and STW IQ estimates explain a large proportion of the variance in obtained IQ scores on the WAIS-III, a large proportion of

Table 3.

Correlations between NART and STW total scores and IQ estimates and WAIS-III IQ scores.

IQ Estimates	WAIS-III IQ Scores: New Zealand European FSIQ	VIQ	PIQ	Māori FSIQ	VIQ	PIQ
NART						
Total Score	.70**	.69**	.53**	.27	.03	.56*
Errors	-.70**	-.70**	-.51**	-.27	-.03	-.56*
FSIQ	.70**	.70**	.51**	.27	.03	.56*
VIQ	.70**	.70**	.51**	.27	.03	.56*
PIQ	.70**	.70**	.51**	.27	.03	.56*
STW						
Total Score	.61**	.60**	.51**	.69**	.32	.85**
FSIQ	.70**	.75**	.47**	.91**	.75**	.59**

* $p < .01$

** $p < .001$

the variance remains unexplained. In order to better understand discrepancies between NART and STW IQ estimates and WAIS-III obtained FSIQ scores, qualitative categorical allocation of individuals by NART and STW estimates to those of obtained WAIS-III scores was examined. All qualitative categories used were in accordance with Wechsler (1997). All of the 14 Māori participants received WAIS-III FSIQ scores which fell into the average range. Of these, 13 (92.9%) received NART FSIQ estimates that fell within the high average range, while the remaining individual received an accurate estimate within the average range. For the STW, 13 (92.9%) of Māori individuals had their FSIQ accurately estimated as falling within the average range, while the remaining individual's FSIQ was overestimated as falling within the high average range. A summary of comparisons for the 75 New Zealand European participants is provided in Table 4.

As can be seen in Table 4, the NART provided accurate FSIQ estimates for 41% of the sample, while the STW provided accurate estimates for 52% of the sample. A ceiling effect is evident for the NART FSIQ estimates, with all participants who had a very superior WAIS-III FSIQ and the majority of those with a superior WAIS-III IQ having their FSIQ underestimated by the NART. The NART was very (88%) accurate in predicting FSIQ scores which fell within the high average range, followed by those falling in the average range (57%). It is notable that all errors in estimation for the average range involved overestimation of FSIQ by the NART. NART accuracy in the upper ranges (very superior and superior) was much less (0% and 8%). In contrast the STW estimates were most accurate within the average range (68%), with accuracy ranging from 39% to 46% across the other ranges presented. Thus, while overall accuracy for the STW exceeded that of the NART, the NART was the best estimator for scores within the high average range for this sample. It is also notable that there was less variability in errors made by the NART than the STW, with any inaccurate estimates made by the NART likely to be within a single category discrepancy from actual WAIS category, whereas STW estimation errors tended to be 1 or 2 categories discrepant, and in one case 3 categories discrepant, from WAIS attained IQ categories.

Discussion

This study examined the NART and STW as predictors of premorbid ability in a sample of healthy New Zealand born adults. New Zealand Europeans produced significantly higher STW FSIQ, and WAIS-III FSIQ, VIQ, and PIQ scores than Māori, but did not perform significantly differently on the NART. These findings are corroborated by those of Halliday (2006) who reports that New Zealand Europeans (n=32) obtained significantly higher scores than Māori (*n* =21) on FSIQ and VIQ scores generated by the Wechsler Abbreviated Intelligence Scale, but found no significant differences in NART performance between the two groups. Unfortunately Halliday reports no subtest score findings with which to

Table 4

Number and percentage of New Zealand Europeans and Maori whose WAIS-III FSIQ fell into each qualitative category and corresponding allocations of NART and STW qualitative categories.

Performances of New Zealand Europeans				
	WAIS-III FSIQ			
FSIQ Estimate	Very Superior (*n* =18)	Superior (n=13)	High Average (*n* = 16)	Average (*n* = 28)
NART				
Very Superior	0	0	0	0
Superior	10 (56%)	**1 (8%)**	0	1 (4%)
High Average	8 (44%)	12 (92%)	**14 (88%)**	11 (39%)
Average	0	0	2 (13%)	**16 (57%)**
STW				
Very Superior	**7 (39%)**	2 (15%)	0	0
Superior	8 (44%)	**6 (46%)**	7 (44%)	6 (21%)
High Average	2 (11%)	3 (23%)	**7 (44%)**	2 (7%)
Average	1 (5.6%)	2 (25%)	2 (13%)	**19 (68%)**
Low Average	0	0	0	1 (4%)
Performances of New Zealand Maori				
	Very Superior (*n* = 0)	Superior (*n* = 0)	High Average (*n* = 0)	Average (*n* = 14)
NART				
High Average	0	0	0	13 (93%)
Average	0	0	0	**1 (7.1%)**
STW				
High Average	0	0	0	1 (7.1%)
Average	0	0	0	**13 (93%)**

Note: Text in bold represents number and proportion of estimates which fell within same category as WAIS-III FSIQ score.

make a comparison. The findings are also similar to those of Ogden, Cooper, and Dudley (2003) who contrasted the performance of 20 New Zealand Europeans and 20 Māori across a number of neuropsychological tests. Of those tests also administered in this study, similar findings are reported in that Māori differed significantly in performance on Vocabulary and did not differ significantly in performance on Digit Span or Digit Symbol Coding. However, Ogden et al.'s Māori and European samples did not differ significantly in STW performance. It is possible that this difference in findings may be related to the different characteristics of the samples, as Ogden et al. compared individuals from low socioeconomic status (SES) groupings, whereas the majority of individuals in the present sample were well educated and worked in professional employment suggesting a much higher SES. That both STW and WAIS performances were impacted by ethnicity is hypothesised to reflect cultural biases common to both measures which were not shared by the NART.

In regards to accuracy of prediction, for New Zealand Europeans the NART total score and STW IQ estimates were most correlated with obtained WAIS-III FSIQ, with each explaining 49% of the variance in FSIQ. This is slightly less than that explained in similar Canadian and English studies, which found that about 59% and 50% of WAIS-R FSIQ short form variance was explained (Sharpe & O'Carroll, 1991; Watt & O'Carroll, 1999), with 66% explained when compared to the full WAIS (Crawford et al., 1989). STW IQ estimates were the most strongly correlated to obtained WAIS-III VIQ, explaining 56% of the variance in VIQ; though it had the weakest relationship to PIQ, explaining only 22% of the variance. This is in contrast to the literature where STW has been found to account for 29% of the variance in VIQ, 13% of variance in FSIQ, and 0.8% of variance in PIQ (Law & O'Carroll, 1998).

For Māori in this sample only the STW FSIQ estimate was significantly correlated with obtained WAIS-III FSIQ, VIQ, and PIQ scores; accounting for 83%, 56%, and 34% of the variance, respectively. These are far in excess of the findings in previous literature, as noted above. These findings are particularly interesting when one considers that those in the Māori sample produced a quite restricted range of IQ scores (falling mainly within the average range), a factor that typically reduces correlations. The strength of the STW as an estimator of premorbid IQ for Māori in the sample was further supported by examination of qualitative labels allocated. All 14 Māori participants received WAIS-III FSIQ estimates which fell into the average range. Of these, 13 (92.9%) received NART FSIQ overestimates that fell within the high average range, while the remaining individual received an accurate estimate within the average range. For STW the opposite pattern was found, with 13 (92.9%) Māori individuals having their FSIQ accurately estimated within the average range, and the remaining individual's FSIQ overestimated as falling within the high average range. This is compelling evidence in favour of use of the STW rather than the NART in estimating premorbid IQ for Māori. In discussing the impact of culture, it must be noted here that Māori performed significantly worse than New Zealand Europeans across a number of WAIS-III subtest scores, all three WAIS-III IQ scores, and the STW IQ estimate. That these differences were not attributable to age, education, gender, history of concussion or current occupation suggests that cultural bias is likely to exist in both of these tests, and that this common source of variance may have contributed to their relationship. It is also noted that while expected relationships were found between education and test performance for New Zealand Europeans, these were absent for Māori. Thus, while clinically it is often assumed that Māori with high levels of achievement within the educational system will perform better on neuropsychological assessments (both due to educational content and greater experiences of being assessed), this was not born out by the data.

Limitations and Future Directions

The sample's overall ability fell within the high average range, was highly educated and relatively young, which limits generalisabilty. The small number in the Māori sample also suggests that caution is needed in applying the findings until they are replicated. Further studies, with larger sample sizes and a greater variety of performance levels are required to determine if the results are indeed generalisable. Keeping cultural considerations in mind, if larger samples were obtained it would be useful to develop and test New Zealand regression equations to be used with the NART and STW when predicting premorbid IQ levels, or if possible, develop a new version of the NART based on New Zealand word-familiarity (Franzen et al., 1997). Despite these limitations, this study is strengthened by its use of the whole WAIS-III battery of tests, rather than just the 'resilient' verbal subtests, or abbreviated versions (e.g., WASI; Wechsler, 1999); and its inclusion of only New Zealand born individuals, which is likely to present a clearer picture of when these findings are most appropriately used.

Conclusion

Overall the NART was not found to be an accurate means of estimating premorbid IQ in this NZ sample. It generally overestimated IQ in average IQ ranges, and underestimated it at higher IQ ranges. This lack of accuracy in estimation was greater than in the standardisation sample, and is very likely due to differences in word familiarity. Given this apparent cultural bias, improving this through adjusting the formulae used for NZ populations, or alternatively, replacing the NART with a NZ equivalent could have a large impact on confidence of clinical findings. While the STW was similarly poor in estimating premorbid IQ in New Zealand Europeans, it produced accurate estimates for over 90% of the small sample of New Zealand Maori included in the study. At present it must be concluded that care needs to be taken when tests based on language familiarity are utilised, with special considerations needed for level of education, IQ levels and cultural diversity.

References

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology, 32(1)*, 55-65.

Baddeley, A., Hazel, E., & Nimmo Smith, I. (1992). The Speed and Capacity of Language-Processing Test. *Suffolk: Thames Valley Test Company*.

Barker-Collo, S. L. (2001). The 60-item Boston Naming Test: *Cultural bias and possible adaptations for New Zealand. Aphasiology, 15(1)*, 85-92.

Barker-Collo, S., Clarkson, A., Cribb, A., & Grogan, M. (2002). The impact of American content on California Verbal Learning Test performance: *A New Zealand illustration. Clinical Neuropsychologist, 16(3)*, 290-299.

Bright, P., Jaldow, E., & Kopelman, 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(6)*, 847-854.

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

Cockburn, J., Keene, J., Hope, T., & Smith, P. (2000). Progressive decline in NART score with increasing dementia severity. *Journal of Clinical and Experimental Neuropsychology, 22(4)*, 508-517.

Collins, L. F. (2000). Comparative validity of NART, NART-R, AMNART, WRAT-III reading, and act scores in predicting WAIS-III FSIQS: Implications for estimation of premorbid intelligence in adults. Collins, Linda F.: U Memphis, US.

Crawford, J. R., Allan, K., Cochrane, R., & Parker, D. (1990a). Assessing the validity of NART-estimated premorbid IQs in the individual case. *British Journal of Clinical Psychology, 29(4)*, 435-436.

Crawford, J .R., Cochrane, R., Besson, J., Parker, D., & Stewart, L. (1990b). Premorbid IQ estimates obtained by combining the NART and demographic variables: Construct validity. *Personality and Individual Differences, 11(2)*, 209-210.

Crawford, J. R., Deary, I., Starr, J., & Whalley, L. (2001). The NART as an index of prior intellectual functioning: A retrospective validity study covering a 66-year interval. *Psychological Medicine, 31(3)*, 451-458.

Crawford, J. R., Parker, D., Stewart, L., Besson, J., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology, 28(3)*, 267-273.

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

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

Freeman, J., Godfrey, H. P., Harris, J. K., & Partridge, F. M. (2001). Utility of demographic equation in detecting impaired NART performance after TBI. *British Journal of Clinical Psychology, 40(2)*, 221-224.

Halliday, T. J. (2006). The development of a New Zealand Adult Reading Test. *Unpublished Master's thesis. University of Waikato, Hamilton, New Zealand*.

Hart, S., Smith, C. M., & Swash, M. (1986). Assessing intellectual deterioration. *British Journal of Clinical Psychology, 25(2)*, 119-124.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47(9)*, 1083-1101.

Klesges, R. C., & Troster, A. I. (1987). A review of premorbid indices of intellectual and neuropsychological functioning: What have we learned in the past five years? *International Journal of Clinical Neuropsychology, 9(1)*, 1-11.

Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. (2004). Some normative and psychometric data for the Geriatric Depression Scale and the Cognitive Failures Questionnaire from a sample of healthy older persons. *New Zealand Journal of Psychology, 33(3)*, 163-170.

Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and Language, 65(3)*, 422-440.

Law , R., & O'Carroll, R. E (1998). Brief Report. A comparison of three measures of estimating premorbid intellectual level in dementia of the Alzheimer's type. *International Journal of Geriatric Psychiatry, 11(10)*, 727-730.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment (4th ed.). New York: Oxford University Press, Inc*.

Morrison, G., Sharkey, V., Allardyce, J., Kelly, R., & McCreadie, R. (2000). Nithsdale Schizophrenia Surveys 21: A longitudinal study of National Adult Reading Test stability. *Psychological Medicine, 30(3)*, 717-720.

Nelson, H. E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social & Clinical Psychology, 14(3)*, 259-267.

Nelson, H. E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex, 14(2)*, 234-244.

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

O'Carroll, R. (1995). The assessment of premorbid ability: *A critical review. Neurocase, 1(1)*, 83-89.

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

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

Paolo, A. M., Troster, A. I., Ryan, J. J., & Koller, W. C. (1997). Comparison of NART and Barona demographic equation premorbid IQ estimates in Alzheimer's disease. *Journal of Clinical Psychology, 53(7)*, 713-722.

Petito, P. A. (2000). The estimation of premorbid intelligence: A comparison of approaches. Petito, Peter Anthony: The Louisiana State U And Agricultural And Mechanical Coll., US.

Riley, G. A., & Simmonds, L. V. (2003). How robust is performance on the National Adult Reading Test following traumatic brain injury? *British Journal of Clinical Psychology, 42(3)*, 319-328.

Saxton, J. A., Ratcliff, G., Dodge, H., Pandav, R., Baddeley, A., & Ganguli, M. (2001). Speed and capacity of language processing test: Normative data from an older American community-dwelling sample. *Applied Neuropsychology, 8(4)*, 193-203.

Schoenberg, M. R., Scott, J. G., Duff, K., & Adams, R. L. (2002). Estimation of WAIS-III intelligence from combined performance and demographic variables: Development of the OPIE-3. *Clinical Neuropsychologist, 16(4)*, 426-438.

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(4)*, 381-384.

Smith, D., Roberts, S., Brewer, W., & Pantelis, C. (1998). Test-retest reliability of the National Adult Reading Test (NART) as an estimate of premorbid IQ in patients with schizophrenia. *Cognitive Neuropsychiatry, 3(1)*, 71-80.

SPSS. (2007) Statistical Package for the Social Sciences for Windows, Release 15.0.1.1. Chicago: SPSS Inc.

Taylor, R. (1999). National Adult Reading Test performance in established dementia. *Archives of Gerontology and Geriatrics, 29(3)*, 291-296.

Thorndike R. L., Hagen E. P., Sattler J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth Edition* (Riverside Publishing Co, Chicago, IL).

Tracy, J. I., McGrory, A. C., Josiassen, R. C., & Monaco, C. A. (1996). A comparison of reading and demographic-based estimates of premorbid intelligence in schizophrenia. *Schizophrenia Research, 22(2)*, 103-109.

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.

Watt, K. J., & O'Carroll, R. E. (1999). Evaluating methods for estimating premorbid intellectual ability in closed head injury. *Journal of Neurology, Neurosurgery & Psychiatry, 66(4)*, 474-479.

Wechsler, D. (1997). WAIS-III administration and scoring manual. San Antonio: The Psychological Corporation.

Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. San Antonio: The Psychological Corporation.

Yuspeh, R. L., & Vanderploeg, R. D. (2000). Spot-the-word: A measure for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology, 15(4)*, 319-326.

Address for Correspondence:

Dr. Suzanne Barker-Collo

Department of Psychology

The University of Auckland

Private Bag 92019

Auckland, New Zealand

Email:
s.barker-collo@auckland.ac.nz

Phone: (64 9) 373 –7599
extension 88517