Psychologists and their administration and scoring of the

NATIONAL ADULT READING TEST (NART)

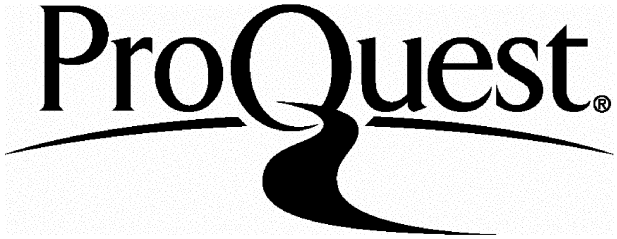
VOLUME I

Darren Mockler

Thesis submitted in partial fulfilment of the requirements

of the DClinPsy

University College London, 1997



VOLUME I

TABLE OF CONTENTS

FIGURES TABLES

TABLES

CHARTS

SCATTTERGRAMS

Abstract

The study investigated the problems psychologists experience when administering and scoring the National Adult Reading Test (restandardised; NART) using the manual pronunciation guideline. Two scoring aids (audio taped version of the NART words and the NART manual pronunciation guideline) were compared to determine which was the most effective in reducing psychologist's NART scoring errors. Three groups of psychologists (Clinical, Trainee and Assistants; N=108) participated in the study. All three groups made a number of errors when pronouncing the NART words. They also experienced difficulty in pronouncing NART word variations. The main finding of the study is that there was no significant difference in NART scoring errors between the two groups of psychologists (using tape or NART manual guideline when scoring) when pre-exposed to either scoring aid. However, a highly significant difference in NART scoring errors was found post-exposure to the scoring aids. The psychologists using the tape as a scoring aid made significantly less scoring errors compared to the psychologists using the NART manual guideline. The standard deviation for the NART scoring errors made by psychologists using the tape does not vary significantly pre- and post-exposure to the tape. However, the standard deviation for the scoring errors made by psychologists using the NART manual guideline significantly increased. The introduction of the manual guideline caused a wider distribution of error scores. An interaction effect was found between the three psychologist groups and the scoring aid used. The findings cast doubt on the current and past use of the NART both for research and clinical practice and the use of the NART manual pronunciation guideline as criteria for scoring word pronunciation.

Acknowledgements

To Dr Valerie Curran, Dr Judith Broadbent and Pasco Fearon for their comments and support. Thank you to all the psychologists, and other study participants for taking the time (and having the courage) to be involved in the study.

Psychologists and their administration and scoring of the NART

Introduction

Estimating premorbid intelligence is an activity central to neuropsychological assessment and research. Early attempts to estimate premorbid intelligence used rather crude methods of locating school records related to educational performance. This approach was superseded by the use of regression equations containing information related to age, sex, race, education and occupation (Wilson et al, 1978; Crawford et al, 1989).

It is well established that a number of demographic variables have a strong relationship with measured IQ (Matarazzo, 1972). Therefore, a patient's educational/occupational history can provide the clinician with a rough guide to their perceived level of premorbid ability. The use of demographic variables for estimating premorbid ability was developed by Wilson et al. (1978), who built regression equations which contained information about age, sex, race, education and occupation. Using the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) standardisation sample (N=1700) Wilson and his colleagues regressed WAIS Full Scale, Verbal and Performance IQ on age, sex, race, education and occupation using a stepwise procedure. Education was the single best predictor of IQ for all the WAIS scales, although all three remaining variables significantly improved predictive accuracy at subsequent steps of analysis. These variables were reported to account for 54%, 53% and 42% of variance in WAIS Full Scale, Verbal and Performance IQ respectively. In a follow up study, the combining of

the demographic estimate of premorbid IQ with a subject's current WAIS performance was found to be of value in discriminating neurological from non-neurological subjects (Wilson et al, 1979). A number of cross-validation studies concluded that Wilson's method is of little utility because, in the samples examined, the equations predicted only a small proportion of IQ variance and systematically overestimated the actual IQ (Bolter et al, 1982; Klesges et al, 1985). However, it was pointed out that the samples consisted of clinically referred subjects (a significant proportion of whom were liable to be cognitively impaired) and are therefore inappropriate for cross-validation purposes (Crawford, 1992a). The more adequately designed cross-validation studies (Karzmak et al, 1985) supported the method developed by Wilson et al, (1978).

In the UK, Crawford et al, (1989a) developed a regression equation (FSIQ=104.12 - 4.38 x class + 0.23 x age + 1.36 x education - 4.7 x sex) that predicted 50%, 50% and 30% of variance in WAIS Full Scale, Verbal and Performance IQ respectively. It was proposed that the demographic method is a reasonable predictor of Full Scale and Verbal IQ in the UK (Crawford, 1992a). However, it can be seen that the proportion of Performance IQ variance predicted was unimpressive for both the demographic methods of Crawford et al (1989) and Wilson et al (1978). Crawford et al, (1990) went on to develop regression equations to estimate premorbid WAIS-R for a UK sample using the procedure employed for the UK WAIS equations. Demographic variables predicted 53%, 53% and 32% of the variance in Full Scale, Verbal and Performance IQ respectively. However, the demographic approach only accounts for

50% of variance in IQ, leaving the remaining 50% unaccounted for. On a positive note, the demographic approach has the advantage of being completely independent of current cognitive functioning.

An alternative historical approach to this problem was the comparison of the 'hold' (eg. the vocabulary subtest of the WAIS) and the 'non-hold' (eg. block design subtest of the WAIS) subtests of the Wechsler scales (Coolidge et al, 1985). This is based on the assumption that certain well learned verbal skills remain relatively unaffected by conditions such as early dementia (Lezak, 1983). However, this approach encountered early problems with even the most highly rated 'hold' tests in the WAIS showing evidence of vulnerability to the effects of dementia type conditions (Hart et al, 1986; Crawford et al, 1988; Sharpe & O'Carroll, 1991).

The National Adult Reading Test (NART; Nelson, 1982) was developed following the observation that reading ability (accuracy of oral pronunciation) was relatively well preserved in individuals with a dementia type illness (Nelson & McKenna, 1975). As such it has become the most widely used estimate of premorbid IQ in both research and clinical practice over the past decade. Prior to the development of the NART, the vocabulary subtest of the Wechsler scales was widely used as a premorbid index (Lezak, 1983). Studies comparing the NART and the vocabulary subtest have consistently showed the former to be the more resistant of the two (Hart et al, 1986; Crawford et al, 1987; Crawford et al, 1988)

The NART is a single short word, oral reading test consisting of 50 items which the subject reads aloud. The use of short words means that subjects do not have to analyse a complex visual array. Also, the words are irregular, this means that they do not follow rules for normal graphemes (the written form of words) - phonemes (a unit of significant sound in a language which differs and distinguishes the two words ie. c from cat differs from b in bat) (e.g. ache, gauche). Intelligent guesswork will not neceassarily provide correct pronunciation; performance therefore depends more on previous knowledge than current cognitive capacity (Nelson & O'Connell, 1978).

Crawford (1992a) reported that the NART IQ correlates significantly with education (r = .51) , social class (r = .36) and age (-.18). The age variable, although significant, accounts for practically none of the IQ variance (Crawford et al, 1988a). No gender effects have been reported (Schlosser & Ivison, 1989). The correlation with demographic variables generally was reported to be .70 (Crawford et al, 1990a). Crawford and colleagues recommended using the NART in conjunction with demographic variables for prediction of premorbid ability in deteriorating patients (Crawford et al, 1990b; Crawford et al, 1990a). These predictions were made from studies using normal subjects. When used with patients experiencing dementia and language disturbances this procedure was reported to underestimate premorbid ability (Stebbins et al, 1990a; Stebbins et al, 1990b). However, the validity problems reported by Stebbins and his colleagues were thought to be associated with the use of a formula

based on data from British subjects tested on the British form of the Wechsler tests for a North American (Canadian) sample (O'Carroll, 1992).

The NART was compared to the demographic method to determine which was the best predictor of premorbid ability (Crawford et al, 1992a). The NART predicted 66%, 72% and 33% of variance in WAIS Full Scale, Verbal and Performance IQ respectively. In comparison, the demographic variables (age, gender, years of education and social class) accounted for 50%, 50% and 30 % of variance in WAIS Full Scale, Verbal and Performance IQ respectively. There is little to choose between both methods for predicting Performance IQ, as both produce fairly disappointing results. However, the NART is significantly superior in predicting Full Scale and Verbal IQ. There is considerable covariance between the NART and demographic variables, most notably in education and social class (Crawford et al, 1988a; Crawford et al, 1990c). For this reason it has been assumed that the combining of these variables with the NART would have no additive affect on predicted IQ variance. However, when the NART was combined with the demographic variables the two methods accounted for 73%, 78% and 39% of variance in WAIS Full Scale, Verbal and Performance IQ respectively. It is evident therefore that a cumulative affect occurred whereby the demographic variables mediated the relationship between the NART and IQ.

Crawford et al (1990b), investigated the construct validity of the combined NART/demographic estimate by factor analysing it in conjunction with the WAIS. The NART/demographic estimate loaded very high on 'g', indeed, its loading (0.9) exceeded that of all the WAIS subtests. A further cross-validation study using the original NART standardisation sample (N=120) was completed (Crawford et al, 1990a) and confirmed that the inclusion of demographic variables significantly increased predicted variance. However, when combining the NART-R with demographic variables (using the WAIS-R as the criterion variable), the demographic variables did not significantly improve predicted variance (Blair & Spreen, 1989; Crawford, 1990).

The NART is reported to be one of the most reliable tests used in clinical practice. It has received extensive investigation prior to and since its initial publication. The NART was found to produce high levels of split half (Nelson, 1982; Crawford et al, 1988a) inter-rater (O'Carroll, 1987; Crawford et al., 1989a) and test-retest reliability (Crawford et al., 1989a). It was reported that subjects scoring for errors produced a split-half reliability coefficient of .90 (Crawford et al, 1988a), inter-rater reliability coefficients between .96 and .98 and test-retest reliability coefficients of .98 (Crawford et al, 1989a). A statistically significant practice effect has been reported but this was of very small magnitude (Crawford et al, 1989a).

Correlations of NART generated Wechsler Adult Intelligence Scales (WAIS and WAIS-R) Full Scale IQ (FSIQ) are in the range of .72 (Nelson, 1982) to .81

(Crawford et al, 1989a). Correlations between NART Verbal IQ (VIQ) and WAIS-R Verbal IQ are very high and whereas in contrast, the NART performance IQ is very low. The NART predicted 72% of variance in WAIS Verbal IQ but only 33% of variance in WAIS Performance IQ (Crawford et al, 1989a).

The NART has been proven to be a valid measure of general intelligence in the normal population (Crawford et al., 1989a). Evidence of the NART's construct validity as a measure of intelligence has been provided by a factor analytic study of the WAIS and NART. In this study the NART loaded very high (0.85) on 'g', the general intelligence factor (Crawford et al., 1989b), which was also present in the analysis of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) subtests. In a factor analytic study combining the NART and the WAIS, a factor identified as Verbal Intelligence was reported in which the NART error score had a high (-.85) loading (Crawford et al, 1989). The NART was found to predict 55%, 60% and 32% of the variance in WAIS full scale, verbal and performance IQ respectively using Nelson's (1982) original standardisation sample (Crawford et al, 1989a).

Crawford and colleagues (1989a) also examined the predictive validity of the NART in a cross-validation sample consisting of 151 healthy subjects who were administered the full length WAIS (in the standardisation sample a seven subtest short form had been used). They reported that NART performance predicted 66%, 72% and 33% of the variance in WAIS full scale, verbal and performance IQ, respectively. It was concluded that the NART has high construct validity as a measure of general ability,

and is a powerful predictor of WAIS full scale and verbal IQ, but is relatively poor at predicting performance IQ (Crawford, 1992a).

Early studies of the NART indicated that reading ability was less affected in dementia than other cognitive measures (Nelson & O'Connell, 1978; Nelson & McKenna, 1975). This finding suggests that the NART has a potentially wide range of applicability in organic and functional disorders for estimating premorbid intelligence. Studies investigating the use of the NART for patients with various forms of dementia (for whom the NART was originally developed) have provided a mixed set of results. Several studies report no apparent detrimental effect on reading ability in Cortical Atrophy (Nelson & O'Connell, 1978) early Alzheimer's Disease, Multi-infarct Dementia (Nebes et al, 1984; O'Carroll & Gilleard, 1986; Cummings et al, 1986; O'Carroll et al, 1987; Crawford et al, 1988; Sharpe, 1990), and Alcoholic Dementia (Crawford et al, 1988). Paque and Warrington (1995) investigated whether reading ability is preserved in Alzheimer's disease or similar degenerative conditions. They reported that although patient performance on the NART declines gradually over time, the deterioration on the WAIS is more rapid and severe. It was evident that the decline in reading across the subject groups was due to those patients who have a lower Verbal IQ than Performance IQ. Paque and Warrington (1995) concluded that generally the NART can be used as a predictor of the premorbid intellectual functioning of a patient with dementia, given that the Verbal IQ is greater than the Performance IQ.

Fromm et al. (1991) assessed patients at yearly intervals over a three year period and showed that Alzheimer's patients scored significantly worse over time. They found that NART scores were significantly correlated with dementia severity in patients at the final stage of testing only, suggesting that NART was sensitive to dementia severity only at the later stages of the disease. These findings are supported by other studies concluding that the NART gave accurate estimates of pre-morbid IQ in the early stages of dementia, but that oral reading performance became impaired in the moderate to severe stages (Hart et al, 1986; Stebbins et al, 1990a). Stebbins and colleagues (1990b) also made the important point that NART performance is compromised in Alzheimer's patients who have language disturbance.

NART performance was found to be highly correlated with severity of illness in Alzheimer's Disease ($r = .56$; severity measured by the mini mental state examination, MMSE) (Patterson et al, 1994). The authors questioned the widespread use of the NART as a premorbid estimate of intelligence, as their results indicate that at a stage of only moderate dementia, the NART underestimates premorbid intelligence by -15 IQ points. They interpret their finding as another example of semantic memory breakdown in Alzheimer's disease (Hodges et al, 1992). The evidence suggests that the NART should only be used in mild Alzheimer's type dementia (ie. somebody scoring >13 on the MMSE; Patterson et al, 1994; O'Carroll et al, 1995).

Other clinical conditions reported to be unaffected by NART performance include Closed Head Injury (Crawford et al, 1988), Depression (Crawford et al, 1987; Austin

et al, 1992), HIV infection (Egan et al, 1990) and Parkinson's Disease (Lees & Smith, 1983; Crawford et al, 1988b). In Depression, Austin and colleagues (1992) reported that although depressed patients performed less well on measures of recall, recognition and psychomotor speed, the group did not differ significantly from their matched controls. Crawford et al. (1987) reported similar findings. One of the most difficult areas of differential diagnosis is dementia versus depression. Investigation into the use of the NART-WAIS-R (Broek & Bradshaw, 1994) and NART-Wechsler Memory Scale (General Memory Quotient; O'Carroll, 1994) discrepancy for estimating decline and differential diagnosis between Dementia and Depression showed that none of the simple neuropsychological analyses using the NART could be recommended for use in clinical practice for the differential diagnosis of dementia and depression.

A number of clinical conditions were found to impair NART performance. Patients with Korsakoff's Syndrome were found to perform less well on the NART than controls individually matched for age, sex and years in full time education (Crawford et al, 1988; O'Carroll et al, 1992a). It was concluded that NART performance is detrimentally affected in Korsakoff's syndrome, and that using the NART to estimate premorbid intelligence level in this condition is invalid. Patients with Huntington's Disease also demonstrated impaired NART performance. In a study comparing six Huntington's patients with individually matched controls the patients performed significantly less well on the NART relative to healthy controls (Crawford et al, 1988). In a more recent study using larger numbers the finding was confirmed (Blackmore et al, 1994). The authors concluded that demographic variables may be a more suitable

method of estimating premorbid intelligence in this clinical group. Alternatively, another proposed test of premorbid ability like the

Spot-The-Word test (Baddeley et al, 1993) may also be of use in these clinical groups. A study investigating the use of the NART in long-term survivors of glioma who had received whole brain prophylactic irradiation found that 16 patients matched with controls (for age, gender, education and social class) made significantly more NART errors than controls, even after controlling for demographic variables. The authors concluded that the NART should be used with caution in survivors of malignant primary brain tumours, particularly if involving the left temporal lobe structures (Ebmeier et al, 1993).

A number of clinical conditions have been shown to impair performance on the NART either totally or in sub-conditions of the illness group. However, clinicians and research investigators continue to use the NART for these clinical groups. To highlight this point the clinical conditions of schizophrenia and Korsakoff's syndrome will be discussed. In schizophrenia three studies have investigated the validity of the NART when used with schizophrenic patients. Crawford et al (1992b), studied two schizophrenic samples (community residents and long-stay residents) together with a healthy sample. All groups were individually matched for age, gender, years of education and social class. NART estimated IQ did not differ significantly when comparing community residents to matched controls. However, NART estimated IQ was significantly lower in the long stay sample. Crawford and his colleagues

concluded that low NART scores in the long stay sample may be a valid estimate of low premorbid IQ. The alternative explanation is that NART performance was impaired by the onset of the disease.

Similar findings were reported from a replication study using some study design modifications (Riordan et al, 1994). Both Crawford et al, (1992b) and Riordan et al (1994) reported that it would be inadvisable to use the NART as a measure of premorbid IQ with such patients. Be that as it may, many investigators continue to use the NART as an estimate of premorbid ability in chronic schizophrenic samples (Dunkley & Rogers, 1994; Jones et al, 1994). However, O'Carroll et al (1992b) reported that NART provides a reasonable estimation of premorbid ability in acutely ill unmedicated schizophrenic patients. A more valid question therefore, is whether a reading test like the NART should be used at all to test premorbid ability in disorders beginning in early life (Jones et al, 1993)?

A similar problem applies to the use of the NART for Korsakoff's Syndrome patients. As previously mentioned NART performance is impaired in this condition. However, like schizophrenia, investigators continue to use the NART as an estimation of premorbid ability for this clinical group (O'Carroll, 1995a).

The use of the NART test for patients who are dyslexic, aphasic, or who have articulatory or visual acuity defects is contra-indicated (Spreen & Strauss, 1991). However, the NART test's sensitivity to the language deterioration in Alzheimer's

disease may make it an effective early predictor of dementia (Schlosser & Ivison, 1989).

A variant of the NART is the shortened version (Beardsall & Brayne, 1990). This uses only the first half of the word list (25 words) to avoid distressing patients who have limited reading skills (the second half of the test is considered to be more difficult than the first half). The test was developed for subjects who fail more than five of the first twenty-five items and are thus confronted with repeated failures. For those who pronounced between 12-20 of these items correctly a full NART score can be estimated. IQ score estimates obtained by this method correlated with full NART estimates with virtually equivalent accuracy. However, these correlations left a considerable unexplained variance (23%-31%) and produced a small number of cases with highly discrepant estimates of ability as defined by the Wechsler IQ scores (Crawford et al, 1991). Crawford and his colleagues concluded that, despite some reservations regarding the shortened NART's practical utility, it can be used in practice to estimate premorbid IQ. However, a more recent study produced less favourable support for the use of the shortened version of the NART in clinical practice. The short form was used on 202 patients consecutively attending a memory clinic. Results indicated that discrepancies between short NART and full NART error scores (four points or more in 59 percent of cases) were outside the bounds of both clinical and statistical acceptability (Bucks et al, 1996). It was concluded that, despite the appeal of a shortened version of the NART to estimate premorbid IQ, without further modification its use in clinical practice would not be relevant.

In 1990, a revised version of Nelson's original NART was produced (Crawford, 1990; NART-R). A number of changes were made to the test itself and to the standardised procedures. Firstly, it was noted that a superior prediction was achieved when the full length WAIS was used as the criterion measure. For this reason a full length WAIS-R was administered to the NART-R standardisation sample. Secondly, although the overall inter-rater reliability of the NART is high, some individual items have been found to provoke low inter-rater agreement rates (eg. the inter-rater agreement for the word *aeon* is nearer to chance than perfect agreement; Crawford et al, 1992a). Furthermore, raters have been found to differ significantly in their strictness/leniency with which they score the NART (Crawford et al, 1989a). For this reason, Crawford replaced 8 of Nelson's initial 50 words (*courteous, catacomb, aeon, puerperal, aver, topiary, prelate, sidereal*) which were found to have low inter-rater reliability with 8 words with high inter-rater reliability (*business, champagne, frigate, orion, spatial, viscera, indict, hyperbole*) (Crawford, 1992a). The NART-R was standardised on a sample of 200 participants recruited to match the UK census in terms of age, sex, and social class distribution (Crawford, 1990). The NART-R predicted 59% of Full Scale IQ variance (Crawford, 1992a).

The NART's popularity in both research and clinical practice for estimating premorbid ability is not limited to the UK. The test has been adopted across the Atlantic in North America and in Europe by the Netherlands. In North America the need for the adaptation of the NART for use in the local population was evident when study findings completed in the United States of America were criticised due to their use of a

test standardised on a UK population with British normative data (Stebbins et al, 1990a; Stebbins et al, 1990b; O'Carroll, 1992). The North American Adult Reading Test (NAART; Blair & Spreen, 1989) was developed to suit both US and Canadian patients. The NAART was found to correlate very highly with the WAIS-R Verbal IQ ($r = .83$). However, the correlation with the WAIS-R Full Scale IQ ($r = .75$) leaves a great deal of variance unaccounted for. The correlation with Performance IQ is not too low ($r = .40$) to be useful as an indicator of premorbid ability. Spreen and Strauss (1991) suggest then that NAART scores work best at the lower limits of estimates of premorbid ability. However, the NAART was found to overestimate IQ scores for normal subjects when the WAIS-R Full Scale IQ was less than 100 and underestimated them when the WAIS-R Full Scale IQ was more than 100 (Wiens et al, 1993). Moreover, the greater the actual IQ score deviation from 100 the more discrepant was the NAART estimate, although the difference between the WAIS-R Full Scale IQ and the NAART estimate was less than 15 points for 95% of the 302 subjects. The problems with adapting a test standardised in one population is clearly evident in the development of the NAART test. Correlations of the NART with the three averaged Wechsler scores were a little lower for an English speaking South African population suggesting that a language test standardised on one population may not work as well with another in which small differences in language have developed over time (Struben & Tredoux, 1989).

In the Netherlands the Dutch Adult Reading Test (DART; Schmand et al, 1991) was developed for use in the national population. The authors reported a high correlation (.85) with Verbal intelligence in healthy controls (N=22) and insensitivity to cerebral deterioration in patients who are brain damaged and patients with dementia (N=53). The test was also reported to be insensitive to cognitive deterioration in a clinical sample with psychosis (N=43).

In the UK the NART test remains the predominant test estimate of premorbid ability. However, more recently a number of other tests have been developed that are used to estimate premorbid intelligence. During the course of their standardisation study on the short NART, Beardsall and Brayne (1990) noticed how a number of the less well-educated subjects mispronounced common words (ie. words that the study participants were likely to use in their everday lives). Beardsall and Huppert (1994) argued that single word reading errors do not necessarily indicate that the subject has no previous familiarity with the word, and proposed that subjects may fail to recognise single words out of context, despite their having a previous lexical entry.

From this initial observation Beardsall and Huppert went on to develop the Cambridge Contextual Reading Test (CCRT; Beardsall & Huppert, 1994) in which the 50 original NART words (from the restandardised test; Nelson & Willison, 1991) are set within sentences to provide a semantic and syntactic context as in every-day life (eg. the lawyer explained that the son who was *heir* to the estate had a large *debt* to pay, because his father's will was *equivocal*). The authors found that both normal subjects

and patients with dementia significantly improved their word pronunciation performance when reading the words in context. This improvement was most noticeable in the patients with dementia and below average readers. Subjects with above average reading ability showed no difference in their NART and CCRT performance. Beardsall and Huppert (1994) concluded therefore that single word reading remains a valid estimate in the above average reading group. The authors claim that the ability to read isolated words becomes compromised in dementia and, as such the use of the NART leads to an invalid underestimate of premorbid IQ. Furthermore, the authors interpreted their findings of improved accuracy of pronunciation of NART words when placed within sentences as a consequence of the effect of context increasing the probability of recognising a stimulus (eg. failure to recognise a known person in an unfamiliar setting). Consequently, by providing the appropriate context, subjects recognise the words as familiar 'thereby accessing the lexicon and the phonological representation of the stored word' (Beardsall & Huppert, 1994; pp. 239-240). The CCRT provides an interesting alternative to the NART and if the claims of the authors are correct a possible replacement for the NART. The CCRT remains only a variant of the NART (ie. using the NART words in sentences).

Although the NART is a valuable test of premorbid intelligence, it does have its limitations. Since it involves reading words aloud it cannot be used with patients suffering from dyslexia, or visual and articulatory problems (Baddeley et al, 1993). It may also underestimate the intelligence of self-educated individuals who may have acquired their knowledge of vocabulary largely through private reading, with the result

that they are familiar with the irregular word, know its meaning, but are unsure about word pronunciation. Consequently, Baddeley et al, (1993) developed a test based on lexical decision making, where subjects are presented with two words, one of which is a real word and one an invented non-word; The subject is then requested to identify the real one. The authors proposed that lexical decision making can be based on any of a number of characteristics of the word including meaning, orthographic appearance, sound or general familiarity. They argued that the presence of a number of parallel routes to perform the task would seem to make it likely that the test would be more resistant to brain damage than a task based on a single feature. The Spot-The -Word test has been shown to be a brief and simple measure of Verbal IQ (Baddeley et al, 1993) due to its high correlation with the NART and Mill Hill Vocabulary tests (.87 and .69 respectively). The relative sensitivity of the Spot-The-Word test to acquired organic impairment has yet to be demonstrated. However, the lexical decision making approach to estimating premorbid IQ utilised in the Spot-The-Word test (a subtest of the Speed and Capacity of Language Processing Test; SCOLP; Baddeley et al, 1990) - has provided a challenge to the NART for usage in estimating premorbid intelligence.

After considering all the approaches to estimating premorbid intelligence it is clearly evident that the NART remains the most widely used estimate of premorbid intelligence and perhaps the test's strongest competitor, the Cambridge Contextual Reading Test, remains only a variation of the original NART test.

Although the NART has been shown to have impressive reliability and validity estimates and is a test which can be used to estimate premorbid ability in various clinical groups it also presents a number of problems associated with its conversion to WAIS-R estimates, what it actually measures, educational, cultural and regional accent biases and problems with how the test is administered and scored.

The NART is often used in close association with the WAIS-R with useful current IQ - estimated IQ discrepancies and NART error conversion tables to WAIS-R Full Scale, Verbal and Performance IQ provided in the NART manual. However, recently Nelson's NART was found to overestimate WAIS-R Full Scale, Verbal and Performance IQ by 5.3, 5.5 and 2.1 respectively in a normal sample (Mockler et al, 1996). In this study the NART (restandardised; Nelson & Willison, 1991) was compared to Crawford's revised NART (Crawford, 1990) to see which test was more comparable to WAIS-R full scale, verbal and performance IQ estimates. The findings demonstrated that Crawford's revised NART was more comparable to WAIS-R on both Full Scale and Verbal IQ. However, both versions of the NART were inadequate estimates of Performance IQ. This raised the question of what cognitive ability are both NARTs measuring? Some studies reported that the NART is a valid measure of general intelligence (Crawford et al, 1989a) in pathological and normal subjects and that the NART loads highly on 'g' (Crawford et al, 1989b). However, the Performance IQ estimates obtained by both NARTs in the Mockler et al (1996) study suggest that the NART and NART-R are not indicative of general intelligence. The NART test may only be measuring the person's reading ability and knowledge of words. If so, are

these abilities adequate to measure verbal intelligence for comparison to the WAIS-R estimates of this function? Given the increasing use of the WAIS-R and NART by neuropsychologists and clinicians, these results have implications for the optimal choice of tests for the estimation of premorbid ability and general intelligence comparisons for estimated and obtained IQ's.

The method generally used for the administration and scoring of the NART test provides a great deal of scope for bias influences and errors in scoring. The assessor would normally administer the test by sitting with the subject and asking the individual to read the words aloud. The assessor would then either rate each word as the person read aloud or make a note of the pronunciation for later scoring (mainly if they are not sure of the pronunciation). For the process of rating the subject's correctness of pronunciation the assessor would need to feel very comfortable with the word pronunciation and possible word variations to rate the patient/study participants. The psychologist would also need to deal with the problems of accent variations at the same time. The problems with accent variations raises another issue: do psychologists make any allowances for accent variation? and if so, what allowances do they make? If psychologists are making allowances for accent variation then this will have obvious implications for the use of normative data and the current standardised method of administering and scoring the NART.

The purpose of the proposed study is not to provide further validation for the NART as an estimate of premorbid intelligence. One of the main areas of interest of the

proposed study is the administration and scoring of NART by Clinical, Trainee and Assistant Psychologists. The assumption is that the NART is used in clinical settings for both assessment and research purposes and that psychologists administering the NART may be making errors when scoring the test.

When administering the NART the test administrator needs to be able to pronounce the words themselves without difficulty and be aware of any variations in word pronunciation. The administration of the NART requires the psychologist to score participant's pronunciation of NART words for errors as they read the words one after the other. This would be a difficult task to complete if the test administrator was not comfortable with the word pronunciations him/herself. Another obstacle is the variety of accents presented by individuals which can lead to difficulty in assessing word pronunciation. Furthermore, there is an assumption that psychologists administering the NART would know the correct pronunciation of the words used. This assumption has no basis in fact. It is possible that many psychologists would not be able to pronounce each NART word correctly (and with variations) and would therefore make errors when scoring the NART even with the assistance of the manual pronunciation guideline provided with the NART.

Another question remains - how does each individual learn the words for the NART before administering the test? The likely answer to this question is either by instruction from another person/supervisor or by the word pronunciation guideline provided with the NART. The first method of learning presents one fundamental problem: if the

supervisor is pronouncing the words incorrectly then their errors would be passed on to the learner. Secondly, the use of the NART manual pronunciation guideline provided with the test does not guarantee understanding of correct word pronunciation. Learning word pronunciation is often facilitated by hearing the correct pronunciation of a word and then spending time practising the use of the word (particularly for those people not schooled extensively in phonetics). One other point is raised concerning the manual pronunciation guideline. A number of inconsistencies exist when comparing various dictionary interpretations of word pronunciation. The inconsistency between dictionaries regarding what is considered correct word pronunciation raises issues concerning the validity of the NART word pronunciation criteria detailed in the NART manual guideline. This point will be explored further in the current study.

The proposed study will investigate the administration of the NART by Clinical, Trainee and Assistant Psychologists who would have preferably used the test in practice, to identify errors in word pronunciation made by psychologists and incorrect scoring of the test. The investigator postulates five possible outcomes following collection and analysis of data. Firstly, Clinical, Trainee and Assistant Psychologist participants will make errors on the NART when the test is administered to each psychologist. Secondly, that Clinical psychologists will make less errors on the NART compared to Trainee and Assistant psychologists. Thirdly, the majority of psychology participants would experience difficulty in pronouncing the NART words correctly and in the use of variation of word pronunciation. Fourthly, a reduction in the number of

NART errors made by psychologists would be associated with the number of times the psychologist had used the NART for research or clinical practice. Fifthly, the experimental group consisting of psychologists using a audio taped version of NART word pronunciations (provided by the English Department at University College London) would make significantly less errors when scoring the NART compared to a control group scoring participant responses to the NART using the NART manual pronunciation guideline.

Method

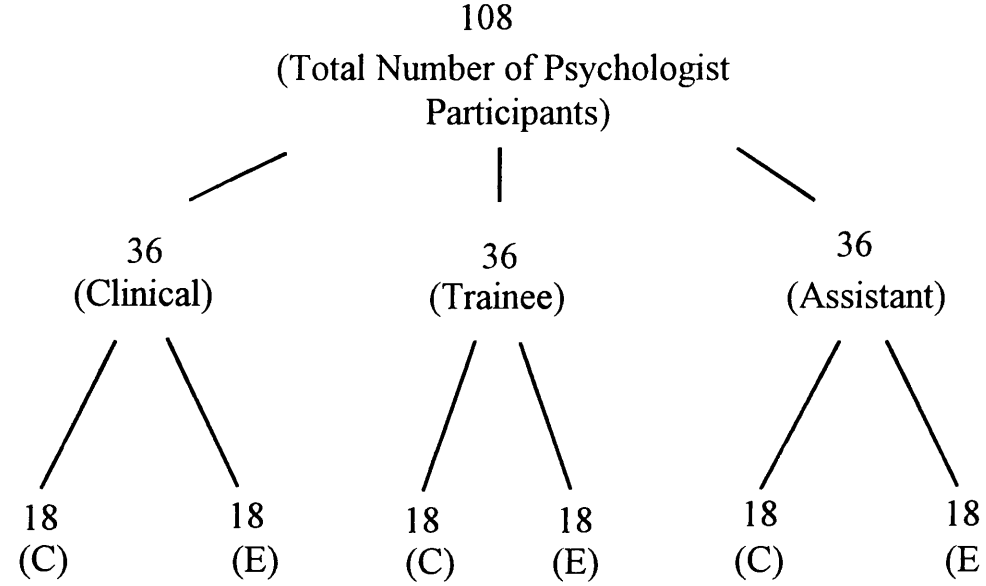
Design

The study is a between subjects design using three groups of psychologists (Qualified Clinical, Trainees and Assistants) (see Diagram 1). The study aims to investigate the effect of two scoring aids (audio taped NART pronunciation guideline and NART manual pronunciation guideline) on NART scoring errors made by the three groups of participating psychologists. The main area of interest is the interaction effect between the three groups of psychologists and the scoring aid used.

Participants

108 participants were recruited from the North and South Thames Region. The participants are representatives of three groups of Psychologists namely: qualified Clinical Psychologists (including neuropsychologists and counselling psychologists), Clinical Psychologists in training and Assistant Psychologists. The psychologists were given the opportunity to ask questions about the study and then asked to sign a consent form prior to participation (see appendix 1). The subjects were predominantly women which is reflective of the male (N=31) to female (N=77) ratio trend existing in the clinical psychology profession (see table 1 for demographic details). The psychologists were recruited from various areas of speciality (adult mental health, older people, forensic, rehabilitation, learning difficulties, and neuropsychology).

Diagrammatic Representation of Research Design
(Diagram 1)



Stage1: Psychologists rate the first set of 5 audio taped participants reading the NART words (pre-exposure to the scoring aid)

Stage 2: Psychologists rate the second set of 5 audio taped participants reading the NART words (post-exposure to the scoring aid)

C = Control Group; Psychologists rating 10 participants using the NART manual pronunciation guideline
E = Experimental Group; Psychologists rating 10 participants using the audio taped version of the NART words

Table1: Demographic details for Psychologist participants

	Clinical	Trainee	Assistant
Number of Psychologists	36	36	36
Mean Age (SD)	39.8 (9.2)	28.8 (4.0)	26.1 (4.1)
Gender (M/F)	13/23	10/26	8/28
MeanTime qualified/ in post (SD)	8.9 (9.2)	2.1 (0.6)	2.0 (1.5)
Speciality:			
Adult Mental Health	12		17
Forensic	4		3
Older People	7		11
Learning Difficulties	2		2
Neuropsychology	8		1
Rehabilitation	3		2

The trainees were not grouped into speciality classes because of their current status of not working in one particular area. Two counselling psychologists were included in the clinical psychologist participant groups.

The criteria for inclusion in the study for the Clinical Psychologists was a British Psychological Society (BPS) recognised qualification including Masters Degree, Practitioner Doctorate (eg DClinPsy) or postgraduate Diploma in Clinical Psychology. To be recognised as a Clinical Neuropsychologist the individual would possess the qualifications of a Clinical Psychologist and would also be employed in clinical practice or research that involves predominantly Neuropsychological practice. Counselling Psychologists were included in the Clinical Psychologist sample if they had used the NART in their clinical or research practice. The Counselling Psychologists would have obtained a Masters Degree in Counselling Psychology recognised by the BPS and be employed in a clinical psychology setting. The Clinical Psychologists in Training recruited for the study were currently enrolled for training leading to the award of the aforementioned qualifications. The Assistant/Research Psychologist would be employed by a Psychology department engaged in Psychology related research or practice and would be of post graduate status. Preferably each psychologist would have administered the NART (restandardised; Nelson & Willison, 1991) for research or clinical practice purposes. However, if the proposed participant had not administered the NART they were given a copy of the manual guideline pronunciation sheet provided in the NART (restandardised) manual and asked to make preparations.

to familiarise themselves with the words and their pronunciation. All of the participants were encouraged to review the NART manual guidelines prior to involvement in the study. The reason for this intervention was to imitate the preparations involved prior to administering the NART to a patient in a clinical situation (ie the psychologist may re-familiarise themselves with the NART word pronunciations prior to completing a psychological assessment).

A further 10 participants were recruited from the general population. The participants were of varied regional and national origin (to provide a diversity of regional accents for interpretation) and all used English as their first language as stated in the NART manual guideline for usage (Nelson, 1982). The effects of accents per se were not of central interest in the study and consequently, the 10 participants were taken mainly from the Greater London area and 2 from Scotland and Ireland (see table 2).

The recruitment of psychologist participants for the study was achieved partly through contact with departmental heads. The heads of department presented information to their team members and then either gave or refused permission to contact members of their department. One problem with this approach was that the study was discussed and minuted at meetings causing a considerable delay in initial contact with possible study participants and the subsequent commencement of data collection. Other methods of recruitment involved attending the assistant psychologist area meetings, presenting the study and circulating information.

Table 2: Demographic details for the 10 general population participants

Subject Number	Age	Gender	Nart Errors	Born
Group A				
1	39	F	8	Hampshire
2	58	M	20	Stepney
3	37	F	25	Scotland
4	49	M	12	Hackney
5	27	F	15	Hampstead
Gp A means	42		16	
Group B				
6	72	F	8	Essex
7	26	M	15	S.Ireland
8	28	F	16	Middlesex
9	40	M	23	NW London
10	34	F	12	Bucks.
Gp B means	40		15	

It is difficult to ascertain how many refusals to participate were associated with presenting the study and circulating information to various groups of psychologists. The response from this approach was very poor with well over 100 information sheets being circulated, resulting in only 5 responses. The most successful method was by direct contact with psychologists and more particularly heads of departments or groups. The number of refusals by contact through departments was 23 qualified clinical psychologists, 2 trainee psychologists and 2 assistant psychologists. Reasons for not wanting to participate in the study were i) lack of time due to work commitments and ii) certain aspects of the study design were considered 'too threatening'. In total there were 27 refusals to participate in the study (excluding subjects contacted through groups and circulars).

Procedure

The 10 participants from the general population were administered the NART following the standardised procedures outlined in the test manual (Nelson & Willison, 1991). The participants were asked to take their time over the word pronunciation in order to determine how they would pronounce the word, then, to say the word and nothing else. The objective of this procedure was to prevent the provision of any additional information which could aid the participating psychologists scoring the taped responses. Any background noise (passing traffic, people talking outside the room) was not removed from tape as this was synonymous with the environmental sounds experienced in a normal working situation.

In addition, when preparing the tape, the 10 participants' NART responses were recorded in rooms and settings similar to the clinical situation in order to prevent any excess of noise interference.

The participants' responses to the NART words were scored using an audio taped version of NART word pronunciations. The audio tape was prepared by a member of the English Department at University College London specialising in English Language usage. The individual recruited to prepare the tape is a post-doctoral lecturer at UCL and also a qualified Speech Therapist. The phonetics for word pronunciation were taken from the English Pronouncing Dictionary (Ramsaran, 1991). The dictionary used is a phonetics dictionary. The dictionary provided the fifty word pronunciations and thirty five extra variations of these words (giving a total of 85 pronunciations). A word pronunciation guide was taken from the phonetics dictionary. This guide was used to complete the audio tape of NART pronunciations. The representative from the English Department read the words aloud from the phonetics guideline sheet and was recorded on to the audio tape.

The 10 audio taped general population participants were rated independently by 3 Psychologists using the same audio taped NART word pronunciation guideline provided by the English Department. The psychologists included a qualified clinical psychologist and two final year clinical psychologists in training. The psychologist ratings of the audio taped participant responses were used to test the reliability of the audio taped version of NART word pronunciation (see the results section for the

outcome of reliability analysis). The audio taped version of the NART words was used to provide a template for scoring the 10 general population participants for NART errors (see appendices 2 for the scoring template for the 10 subjects). The 10 participants were to be rated for NART errors by participating psychologists in stage 2 of the study. The audio taped version of the NART word pronunciations was also used for rating NART errors made by the three groups of psychologists (Clinical, Trainee and Assistant Psychologists) in stage 1 of the study (see below). The 10 participants were divided into two groups of five (A & B) to be scored for NART errors by the participating psychologists. To 'balance' the two groups the allocation was broadly based on NART error score, gender and place of birth (see table 2).

Stage 1:

The standard version of the NART (restandardised) was administered to each participating psychologist. A variation to standard procedures was used whereby each participant was asked to provide any correct variations they knew to the pronunciation of each word. The NART word booklet is provided for the participant. The following instructions are given:

"I want you to read slowly down the list of words starting here". Indicate CHORD.
"After each word please wait until I say next before reading the next word".

(Nelson, 1982 pg. 2; Ache was replaced by Chord because Chord is the first word in the word booklet for the restandardised NART)

"I would also like you to give any correct variations in pronunciation for any particular word".

The psychologists read the words from the NART word booklet. The psychologists NART pronunciation responses were audio-taped. The taped responses of the participants were then compared to the audio-taped NART word pronunciation guideline - including all variations of correct word pronunciation for each word - prepared by the English Department at University College London. The comparison was used to aid the scoring of participating psychologists NART word pronunciations for errors.

Stage 2:

The 3 participant groups - Clinical, Trainee & Assistant Psychologists - were first compared for NART error scores. The psychology subjects were then divided into 2 groups:

Group 1 (Experimental)

This group consists of 18 participants from each of the Clinical, Trainee & Assistant Psychologist groups (total N = 54).

* The psychologists were asked to rate five (set A or B) of the participant audio-taped NART responses for errors without the use of any scoring aid.

* After rating the first five subjects the psychologists were given the audio-taped NART word pronunciation guideline. The psychologists were allowed a maximum of 15 minutes to familiarise themselves with the taped version of the NART words. The psychologists were told to use the time to prepare themselves before scoring the next five subjects. The participants were allowed to discontinue their preparation for scoring the next five subjects before 15 minutes had elapsed. After this time the audio taped NART word pronunciation guideline was removed.

* The psychologists were then asked to rate the remaining 5 participant audio-taped NART word responses for errors.

Group 2 (Control)

This group consists of 18 participants from each of the Clinical, Trainee & Assistant Psychologists groups (total N = 54).

* The psychologists were asked to rate five (set A or B) of the participant audio taped NART responses for errors without the use of any scoring aid.

* After rating the first five subjects the psychologists were given the NART (restandardised) manual guideline (Nelson & Willison, 1991) for word pronunciation. The psychologists were allowed a maximum of 15 minutes to familiarise themselves with the NART words using the guideline. They were told to use the time to prepare themselves before scoring the next five subjects. The participants were allowed to discontinue their preparation for scoring the next five subjects before 15 minutes had elapsed. After this time the NART manual pronunciation guideline was removed.

* The psychologists were then asked to rate the remaining 5 participant audio taped NART word responses for errors.

The order in which set A and B tapes were given was counterbalanced across participants and design. Thus set A then set B was presented for 50% of participants

in each of the three groups; set B then set A for the other 50% of participants in each of the three groups. This aimed to control version/order effects.

The psychologists from both the Experimental and Control groups' rating of the 10 participant audio taped responses would then be compared for differences in NART error scoring.

Other variables investigated in the study include: (i) the number of times psychologists had previously administered the NART, (ii) the time period lapsed between psychologist's participation in the study and the last time they used the NART test in practice, (iii) the association between NART word errors/NART word variation errors (made by psychologists on the NART test at stage 1 of the study) and the NART scoring errors made by psychologists pre- and post- exposure to either scoring aid (at stage 2 of the study), (iv) and what factors (if any) influenced psychologists' scoring of the NART and psychologists' preference in scoring/learning aids for word pronunciation.

Scoring

Five scores were obtained from the two stages of the study. From stage 1 three scores were recorded:

* NART ERROR SCORE: for each psychologist on the 50 words using the restandardised test.

* WORD VARIATION ERROR SCORE: this score consisted of the number of NART word variation errors made by psychologists in attempting to pronounce the combined 50 words and other possible word variations (total number = 57) presented in the restandardised NART pronunciation guide.

* INCORRECT WORD VARIATIONS: This score consisted of NART word variations given by participating psychologists which they considered to be correct variations of NART words but were actually incorrect word variants

(see appendices 3 for example of score sheets used for obtaining all 3 scores in stage 1) Stage 2 produced the remaining two scores used in the data analysis:

* TOTAL COMBINED NART ERROR SCORE (A): This is the combined total of NART scoring errors made by each psychologist when rating the first set of five general population participants' pre-exposure to either scoring aid.

* TOTAL COMBINED NART ERROR SCORE (B): This is the combined total of NART scoring errors made by each psychologist when rating the second set of five general population participants following exposure to either one of the scoring aids (tape/manual guideline) for 15 minutes.

(see appendix 4 for example of score sheets used for obtaining all 3 scores in stage 2)

A possible advantage was identified for the psychologists using the audio taped NART word pronunciation guide compared to psychologists using the NART manual pronunciation guideline. The NART words used on the audio tape were taken from a phonetics dictionary containing 85 possible variations of the NART words (see table 2a). The words used in the NART manual guideline were originally taken from the Chambers Dictionary providing only 57 NART word variations. The audio taped NART words were taken from a phonetics dictionary (Ramsaran, 1991) in preference to the Chambers Dictionary. This was due to reservations expressed by the representative of the English Department at UCL regarding the insufficient and perhaps invalid current NART manual pronunciation guideline criteria. She proposed that the phonetics dictionary criteria for NART word pronunciation was a more appropriate guideline.

As the tape was to provide the template for scoring the 10 general population participants for NART errors rated by psychologists in stage 2 of the study, a check for any advantage needed to be made. It was necessary to check how many word pronunciations made by the 10 participants could be rated differently (obtaining a correct score) by psychologists using the tape, that psychologists using the NART manual pronunciation guideline would not be aware of (and consequently would provide an incorrect score). Following investigation of the 10 taped participants (by checking which subjects pronounced words in a way that would be scored correct using the tape but incorrect using the manual guideline) psychologists given tape A first would be able to make 3 less errors out of a possible 250 error scores.

Psychologists scoring set B would be able to make 5 less errors out of a possible 250 error scores. Consequently, depending on whether the psychologist was rating the A or B group second, after exposure to the tape, the psychologists would have a possible 3 or 5 point advantage in rating error scores if they were able to learn from the tape and score these words correctly. The balance of order helps here (tape > manual = -3 less errors A vs B; tape > manual = -3-5 less errors B vs A), leaving the advantage similar over the order issue. The difference would only be considered if the mean difference in change in error scoring by psychologists on the 10 participants was very low (ie. a minor level of statistical significance).

Statistical Analysis

The Cohen's Kappa test was used to obtain an adjusted inter-rater reliability estimate of the agreement between 3 psychologists rating the 10 general population audio taped participants for NART errors using the audio taped NART word pronunciation guide. Each pair of NART error scores (psychologists 1-2; 1-3, 2-3; rating each word for the 10 participants; total number of words rated = 500) were calculated individually to produce a percentage agreement range. The scores were then adjusted to account for possible chance agreement. The analysis was completed to estimate the inter-rater reliability. This was necessary to determine whether the rating would be significant enough to use the audio-taped NART pronunciation guideline as a learning/scoring aid to rate psychologists error scoring in stage 1 and to develop a template for rating

NART scoring errors (for the 10 general population participants) made by psychologists in stage 2 of the study.

For data analysis of stage 1 of the study a number of inferential statistical tests were used. The main test implemented was a one-way Analysis of Variance (ANOVA) for a between groups comparison. The test was used to compare the three groups of psychologists (Clinical, Trainee, Assistants) for *NART errors* on the 50 words using the restandardised test. ANOVA was also used to compare *word variation errors* made by participating psychologists. This score was obtained from the number of errors psychologists made in attempting to pronounce the combined 50 words and other possible word variations (total number = 57). The remaining data collected in the first stage from the psychologists is the *incorrect word variations score.* This score consisted of NART word variations given by participating psychologists which they considered to be correct variations of NART words but were actually incorrect word variants. This score was analysed using one-way ANOVA and descriptive statistics. The standard NART error score (out of 50) was also correlated with the number of times psychologists had used the NART test using Pearson's product-moment correlation. These correlation's were calculated using all 108 psychologist subjects in one group to increase variance and statistical power.

The data collected at stage two of the study was analysed using ANCOVA (analysis of covariance) (a between subjects design). Prior to using Analysis of Variance to explore group differences in NART scoring errors made by participating psychologists

regression analysis was applied to identify any covariates used when analysing NART scoring errors in stage 2. The ANCOVA test was used to compare *total combined NART scoring errors (pre-exposure to the scoring aids; tape or manual guideline)* made by the three groups of psychologists scoring the 10 participants (between subjects - the three groups of psychologists).

Using Pearson's Product Moment correlation method the total combined NART error scores were also investigated for associations with the frequency/recency of NART test usage, the order of scoring the two sets of five general population participants, psychologists' self-reported influences when scoring NART errors and NART errors made by psychologists at stage 1 of the study. This was calculated using all 108 psychologist subjects in one group to increase the level of statistical power. The investigation of a possible significant association between NART errors made by psychologists, their frequency/recency of using the NART test and NART scoring errors when rating the 10 general population participants was completed to determine whether any of these factors would need to be used as covariates when analysing the total combined error scores made by the psychologists rating the 10 participants.

The *total combined NART scoring errors (post-exposure to scoring either aids)* made by each psychologist for the second set of five participants was analysed using ANCOVA (analysis of covariance) methods. The ANCOVA was used to compare the between group effects (three groups of psychologists) and the within-group effects (A vs B) on *NART scoring errors.*

When comparing the difference in NART scoring errors within groups for differences in errors for A (pre-exposure to learning aid) and B (post-exposure to learning aid) the participating psychologists groups were investigated as a combined group (ie. 54 psychologists in each group) to increase statistical power.

The ANCOVAs used provided an interaction effect on NART scoring errors made by psychologists: A: pre-exposure to learning aid and B: post exposure to learning aid (ie A vs B x tape vs manual guideline)

A simple t-test was also used to show the difference in NART scoring errors made by the 3 groups of psychologists (using the tape or manual guideline as a scoring aid) rating the first 5 participants (A) and then the second 5 participants (B). The assumption was that the NART scoring errors made by psychologists rating the first five participants (A) should be comparable between groups as no aid to scoring would have been introduced at this point. However, the difference in NART scoring errors made by psychologists between the two groups scoring the second set of five participants (B) should be significantly different following the introduction of either learning aid.

Results

In addition to the main NART data collected, a large amount of demographic and other information was also obtained from participating psychologists. This involved a number of variables assumed to be associated with psychologists and their performance at stage 1 (pronouncing the NART words) and stage 2 (scoring NART errors for 10 general population audio taped participants) of the study. The variables include: (i) the number of times psychologists had administered the NART test, identified biases affecting psychologists' scoring of NART errors, (ii) the time lapsed prior to the psychologists' last administration of the NART test in practice, (iii) the order in which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first) and (iv) their method of learning how to pronounce the NART words prior to using the test. Each variable will be presented individually.

The number of times psychologists had administered the NART test (see chart 1)

The participating psychologists were a mixture of frequent and less frequent users of the NART test in practice. It was important to get a mix of NART users in the study in order to make any generalisations about the findings. The number of times each psychologist used the test was based on the whole of their careers. Therefore, the spread of usage may be over a long period of time. Fractionally under 55% of the participating psychologists had used the NART test on less than 11 occasions or not at

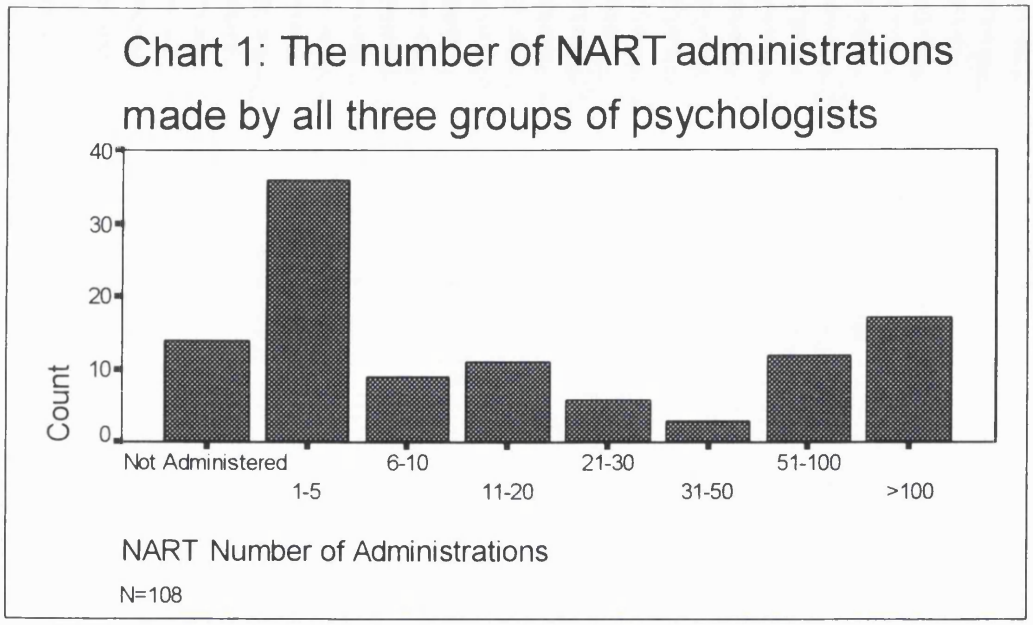


Chart 1: The number of NART administrations made by all three groups of psychologists
Count
40
30
20
10
0
Not Administered
1-5
6-10
11-20
21-30
31-50
51-100
>100
NART Number of Administrations
N=108

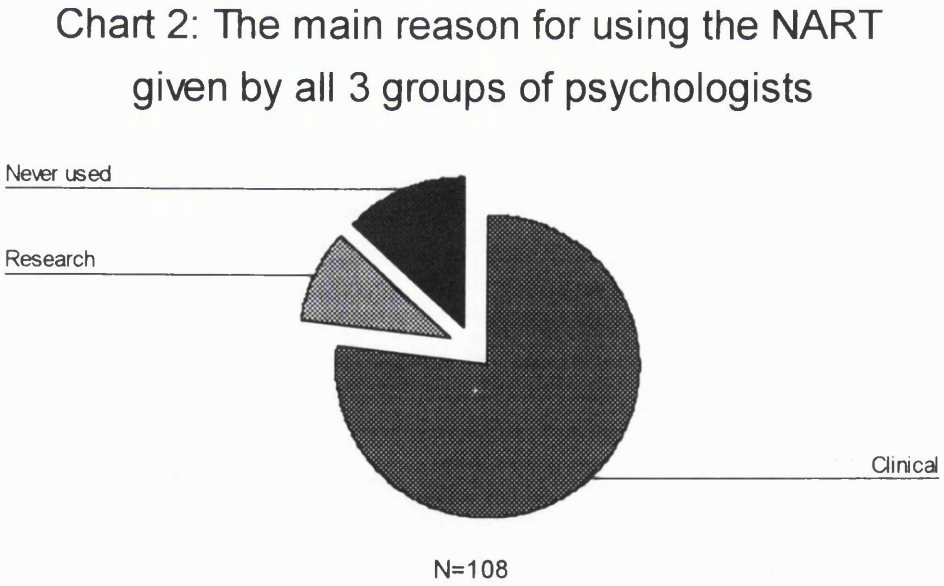


Chart 2: The main reason for using the NART given by all 3 groups of psychologists
Never used
Research
Clinical
N=108

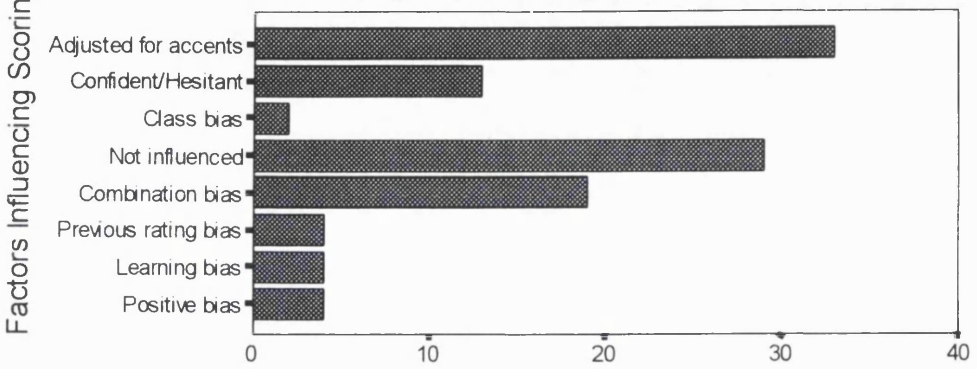
all. At the other end of the scale 14% of psychologists had used the test between 50-100 times and 16% on more than 100 occasions. The main reason (77%) given by participating psychologists for using the NART was for clinical reasons (ie. clinical assessment) and research purposes (10%) (see chart 2).

Factors identified by psychologists affecting scoring of NART errors (see chart 3)

The factors reported by psychologists were based on their experience of scoring the 10 general population participants as part of the study and their general experience using the test in practice. Each factor was coded so the information may be analysed using both descriptive and inferential statistics. A number of factors were identified namely:

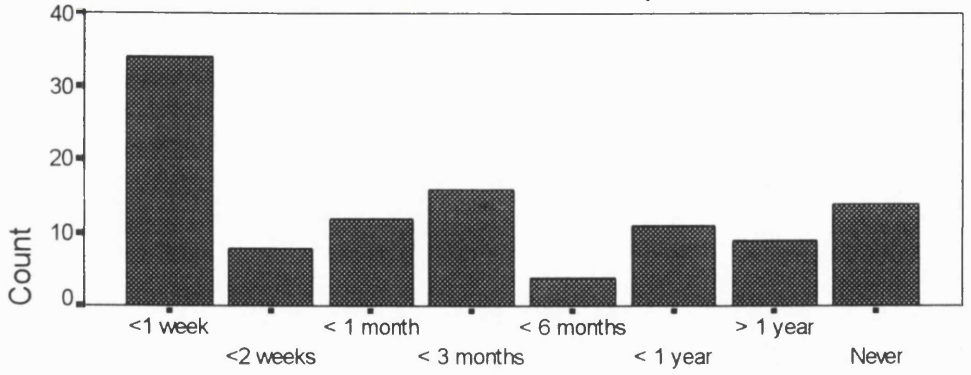
Adjustment for accents: When psychologists attempted to adjust their scoring of NART errors to facilitate a patient/study participant who presented with a perceived accent often different from their own (North of England, East End of London, Scottish, Welsh, Irish etc). Psychologists often attempted to place the word pronunciation in the context of their region (eg. is that how they would pronounce the word in that area?). The result of this was often positive with the psychologist usually giving the person the *benefit of the doubt* and scoring them correctly. However, psychologists were more stringent in their scoring of people with familiar accents (or similar to their own). The accent adjustments made by psychologists accounted for 31% of the factors influencing their decision to score a person right or wrong on the NART.

Chart 3: Factors influencing scoring of errors by all three groups of psychologis



N=108

Chart 4: The time lapsed since each psychologist last administered the NART in practice



N=108

Confident/Hesitant: Psychologists reported that they would often be influenced by the confidence or lack of confidence in the way the person pronounced the word. If the person was very confident some psychologists would be more inclined to score them correct. If on the other hand the person was hesitant - prior to the delivery of the word - this sent a message to the psychologist that they were not sure of the correct pronunciation and would often result in a incorrect score. The confidence/hesitant factor seemed to be more prevalent in psychologists who were less familiar with the NART words and in some cases were unsure of the word pronunciation themselves. This factor was reported by 12% of the participating psychologists.

Previous rating, learning, positive and class influences: The *previous rating bias* involved a positive/negative approach to scoring by psychologists based on the patient/study participants' earlier performance on the NART test. If the person scored poorly in the early section of the test the psychologists were less likely to give them the benefit of the doubt in the latter section of the test. The psychologists also found it difficult to deal with a correct word pronunciation made in the latter stages of the test by a person who scored a number of incorrect word pronunciations in the earlier part of the test. Psychologists seemed to be anticipating an incorrect score based on their earlier performance. This worked in reverse for high scorers with psychologists adopting a more positive bias to scoring (ie. if the person scored highly in the earlier section of the test the psychologists were more likely to give them the benefit of the doubt in the latter section of the test, often anticipating a correct response).

The *learning bias* concerns a change in the psychologist's view of whether a word pronunciation is correct in favour of how the patient/study participants (that

psychologists are rating for NART errors) pronounce the word (ie. the psychologist replaces their interpretation of the correct pronunciation with the subject's interpretation). This generally occurred in the study when psychologists were rating 10 participants consecutively.

The *positive bias* involved the psychologist using a general positive approach to rating NART errors by giving the patient/study participant a correct score if in any doubt about their word pronunciation. The *class bias* was associated with the psychologists view that the patient/study participant was *well spoken* (more likely to score correct) or *not very well spoken* (more likely to score incorrect). This factor was more associated with psychologists who were less familiar with the NART words and their pronunciation. These four factors were reported by 13% of the participating psychologists.

Combination of Biases when scoring: In a number of cases psychologists reported more than one perceived influence on their scoring of NART errors. This factor involved mainly accent, and confidence biases and to a lesser extent a previous rating bias. The factors are described in exactly the same way as reported for each variable previously. The combination of influences was reported by 18% of the participating psychologists. If this figure is combined with the main accent and confidence factors it accounts for 60% of influences reported by psychologists.

Psychologists reporting no influence: It is important to conclude that not all psychologists reported influences affecting their scoring of NART word pronunciation. 27% of participating psychologists were not aware of any influences affecting their

scoring of NART errors when administering the test both in the study and their clinical practice.

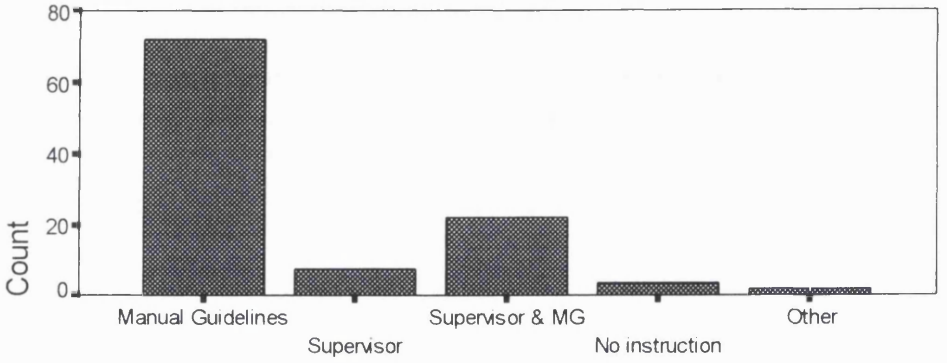
The time lapsed since last administration of the NART test (see chart 4)

The participating psychologists were generally recent users of the NART test in practice. 50% of psychologists had used the test in the last month prior to participation in the study and just under 65% in under 3 months. The largest group of psychologists had last used the test in the previous week. However, 8% of psychologists had not used the NART test in over a year and 13% had never used the test previously. All the psychologists, whether frequent or infrequent users of the test, were encouraged to use the NART manual guideline to practice word pronunciation prior to participation in the study. This was considered an important factor and one that would probably affect psychologist performance in the study. This variable is considered in more detail in a later part of the results section.

The method of learning NART word pronunciation (see chart 5)

In the majority of cases psychologists had used either the NART manual guideline (67% of participating psychologists) or a combination of the manual guideline and instruction from a supervisor (20%) to learn the correct NART word pronunciations. The remaining psychologists used either instruction from a supervisor (7%), no instruction (4%) or other (2%). 'Other' consisted of a discussion with colleagues about

Chart 5: Method of learning NART word pronunciation by all three groups of psychologist



LEARN

N=108

what is considered the correct word pronunciations. The main method of learning the NART word pronunciations was by using the NART manual guideline and/or instruction from a supervisor. The combined reported use of these two methods by participating psychologists equalled 94%.

Inter-rater reliability of the NART audio tape pronunciation guide

The Cohen's Kappa test was used to obtain an adjusted inter-rater reliability estimate of the agreement between 3 psychologists rating the 10 general population audio taped participants for NART errors using the audio taped NART word pronunciation guide. Each pair of NART error scores (psychologists 1-2; 1-3, 2-3; rating each word for the 10 participants; total number of words rated = 500) were calculated individually to produce a percentage agreement range. The scores were then adjusted to account for possible chance agreement. The percentage agreement between the three psychologist raters of NART errors for the 10 general population participants equalled 95-96% agreement. After adjusting for chance agreement using Cohen's Kappa (Cohen, 1960) the level of agreement was reduced to a Kappa of .88. This can be defined as a good level of agreement (Barker et al, 1994). Consequently, the tape was deemed to be a reliable aid for rating the errors made by psychologists when completing the NART test themselves in stage one and for use in providing a template for scoring the 10 general population participants for NART errors.

Stage 1 analysis

Prior to stage 1 and stage 2 the data to be analysed was checked for homogeneity of variance using Levene's test (1960). The data was not heterogeneous and therefore parametric tests were used for all forms of data analysis.

A number of factors were considered in stage 1 of the data analysis. The Psychologists were divided into three groups for data analysis (see table 1 in the methods section for demographic details). The main area of interest at this stage was the difference in NART errors (out of 50), NART word variation errors (out of 57) and incorrect NART word variations made between the three groups of Psychologists. A one-way ANOVA (between-subjects design) was used to identify the overall group differences in errors made by psychologists on the NART. A significant difference was found between the three groups for NART errors (scoring out of 50 words) ($F(2,105) = 17.8$, $p<.00001$) and for NART word variation errors (scoring out of 57 words) ($F(2, 105) = 16.6$, $p<.00001$). *A priori* comparisons were chosen to investigate more closely the differences in NART error scoring between the three groups. The *a priori* test was chosen in preference to the *post hoc* comparisons test because the comparisons were planned prior to the data collection (Howells, 1992; pg 341). The *a priori* test used for between-group comparisons is the Modified Least Significant Difference (or Bonferroni) test. The significance level was set at .05. The Assistant Psychologists made significantly more NART errors and NART word variation errors

than both the Clinical and Trainee Psychologists. There was no significant difference in NART errors between the Trainee and Clinical Psychologists. However, there was a significant difference in NART word variation errors between Trainee and Clinical Psychologists with Trainees making more errors (see table 3 and table 4 for descriptive statistics).

The third score of interest obtained at stage 1 was the incorrect NART word variations given by participating psychologists. This score consisted of NART word variations which participating psychologists considered to be correct variations of NART words but were in actual fact, incorrect word variants. The psychologists were asked to provide any correct NART word variations at stage 1 of the study. A number of the psychologists did not attempt to give any word variations (N=30; 29%) and only provided a standard single pronunciation for each word (50 pronunciations). There was no significant difference between the three psychology groups in number of incorrect NART word variations ($F(2, 105) = 0.13$, $p<.882$) (see table 5).

The psychologists were then allocated to two groups for stage 2 data analysis. The two groups consisted of i) psychologists that would be given the audio-taped version of the NART as a learning aid (experimental; N=54) ii) and psychologists that would be given the NART manual guideline as a learning aid (control; N=54). The two groups comprised Clinical, Trainee and Assistant Psychologists (18 of each in both groups). The allocation of psychologists was based on their NART error scores obtained at stage 1, balanced as such that there was no significant between group

Table 3: The mean number of NART errors made by the three Psychologist groups

	Clinical	Trainee	Assistant
Number of Psychologists	36	36	36
Mean NART Errors	4.8	6.9	10.9
Standard Deviation	4	4.7	4.6
Standard Error	0.7	0.8	0.8
Minimum/Maximum	0/16	1/22	3/22

Table 4: The mean number of NART variation errors made by the three Psychologist groups

	Clinical	Trainee	Assistant
Number of Psychologists	36	36	36
Mean NART Variation Errors	10.8	14.1	17.3
Standard Deviation	4.5	5.1	4.7
Standard Error	0.8	0.9	0.8
Minimum/Maximum	5/23	7/32	8/29

differences ($F(2, 106) = 0.04$, $df = 1$, $p<.837$; see table 6 for descriptive statistics) in psychologists NART errors. This balancing prevented any score bias affecting the evaluation of the scoring aids introduced at stage 2. This allowed for a fair comparison to be made between the psychologist groups and each scoring aid. The balancing of groups on the basis of NART errors made by psychologists at stage 1 was supported by the significantly high correlation between NART errors made by psychologists at stage 1 and errors made in scoring the 10 general population participants at stage 2 pre- ($r = .83$, $p<.0001$) and post- ($r = .73$, $p<.0001$) introduction of the scoring aids (tape and manual guideline).

Stage 2 analysis

Correlation analysis

Prior to analysing the differences in psychologists' NART error scoring for the two groups of five general population participants it was necessary to use correlation analyses to identify any associations between various factors which may affect the interpretation of differences in scoring NART errors. These factors need to be identified before further analysis can take place. The factors investigated at this stage involved the association between the total number of errors made by psychologists scoring the first five subjects (prior to introduction of either learning aid) and its relationship with: (i) the number of times the psychologists had administered the NART test ($r = -.51$, $p<.0001$; inverse correlation); (ii) self reported biases influencing

Table 5: The mean number of incorrect NART word variations made by the three Psychologist groups

	Clinical	Trainee	Assistant
Number of psychologists	36	36	36
Incorrect NART Word Variations	1.6	1.8	1.7
Standard Deviation	1.7	1.7	1.6
Standard Error	0.3	0.3	0.3
Minimum/Maximum	0/6	0/7	0/7

Table 6: The mean number of NART Errors/Variation errors made by the two Psychologist groups divided by aid to be used in stage 2 (tape/manual guideline)

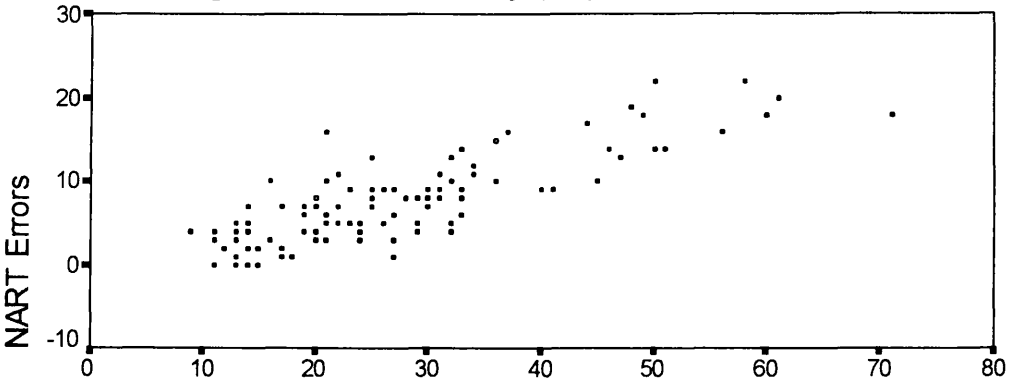
	Tape	Manual Guideline
Number of Psychologists	54	54
Mean NART Errors (Variation Errors)	7.4 (13.9)	7.6 (14.3)
Standard Deviation (Variation SD)	5.3 (5.9)	4.9 (5.0)
Standard Error (Variation SE)	0.7 (0.8)	0.7 (0.7)
Minimum/Maximum (Variation min/max)	0/22 (5/32)	0/22 (6/29)

psychologists' scoring of NART errors ($r = -.02$, $p<.410$, NS); (iii) the time lapsed prior to the psychologists' last administration of the NART test in practice ($r = .39$, $p<.0001$; positive correlation); (iv) the order with which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first) ($r = .30$, $p<.001$; positive correlation); (v) the number of NART errors ($r = .83$, $p<.0000$; positive correlation) (vi) and NART word variation errors ($r = .76$, $p<.0000$; positive correlation) psychologists had made on the NART test at stage 1.

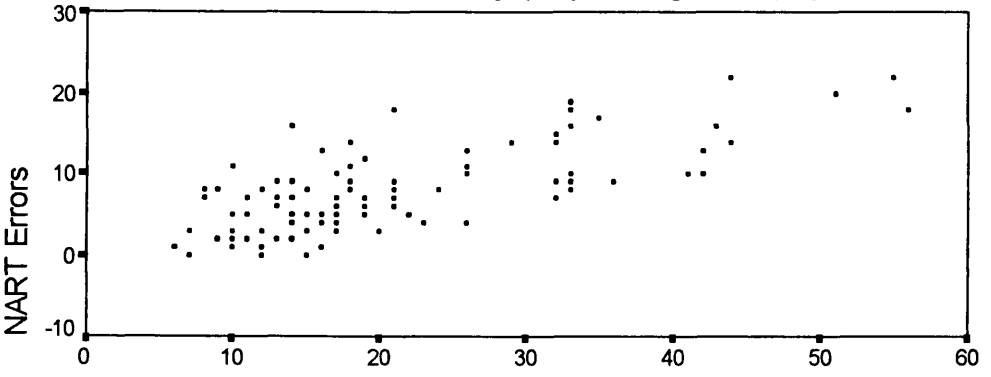
Five of the six variables investigated were found to have a significant association with psychologists' NART scoring errors rating the first five subjects prior to exposure to the scoring aids (tape or manual guideline). The self reported biases influencing psychologists' scoring of NART errors was the only non-significant factor. The five remaining factors were analysed in more detail using regression analysis (see next section) to identify the level of involvement each variable had on psychologists' NART error scoring of the first set of five subjects. The most significant variable was the NART errors psychologists made on the test at stage 1 of the study (see scattergram 1).

The same correlational analysis was completed on the second set of NART error score ratings made by psychologists on the five general population participants after exposure to either scoring aid (audio tape or manual guideline pronunciation guide). The factors investigated at this stage involved the association between the total number of errors made by psychologists scoring the second five subjects and its relationship

S1: Association between NART errors and NART scoring errors made by psychologists (Gp1)



S2: Association between NART errors and NART scoring errors made by psycholgists (Gp2)



with: (i) the exposure to type of scoring aid ($r = .39$, $p<.0001$; positive correlation); (ii) the number of times the psychologists had administered the NART test ($r = -.36$, $p<.0001$; inverse correlation); (iii) self reported biases influencing psychologists' scoring of NART errors ($r = -.009$, $p<.464$; NS); (iv) the time lapsed prior to the psychologists' last administration of the NART test in practice ($r = .24$, $p<.006$; positive correlation); (v) the order with which the subjects rated the 10 general population participants (ie 1-5 first or 6-10 first) ($r = -.003$, $p<.490$; NS); (vi) and the number of NART errors ($r = .73$, $p<.00001$; positive correlation) and NART word variation errors ($r = .70$, $p<.00001$; positive correlation) psychologists had made themselves at stage 1.

Five of the seven variables investigated were found to have a significant association with psychologists' NART scoring errors rating the first five subjects prior to exposure to the scoring aids (tape or manual guideline). The identified biases affecting psychologists' scoring of NART errors and the order with which the subjects rated the 10 general population participants variable are the only two non-significant factors. The five remaining factors were analysed in more detail using regression analysis (see next section) to identify the level of influence each variable had on psychologists' NART error scoring post-exposure to either scoring aid. The most highly significant variable was the NART errors psychologists made on the test at stage 1 of the study (see scattergram 2).

Regression analysis

A multiple regression analysis was completed due to the significant correlations found between the total number of errors made by psychologists scoring the first five subjects (prior to introduction of either learning aid) and the number of times the psychologists had administered the NART test, the time lapsed prior to the psychologists last administration of the NART test in practice, the order with which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first) and the number of NART errors and NART word variation errors psychologists had made themselves at stage 1. The regression analysis was administered to decide which variables would need to be used as covariates when analysing the differences in the three psychologist groups' NART error score rating of the 10 participants pre- and post-implementation of each scoring aid (tape or manual guideline).

The regression analysis produced highly significant results. A stepwise regression was used with the PIN set at .001 and the POUT set at .05. When accounting for the variance in errors psychologists made when scoring the first five participants (Gp1) the only variable to enter the equation at the .001 level was the number of errors psychologists made in stage one of the study (when completing the NART themselves) ($F(1, 106) = 230.5$, $p<.000001$). This accounted for 68% of variance in the NART scoring errors for Gp1 made by psychologists (using the adjusted R square statistic). Two other variables that would have entered the equation if the PIN had been set at .05 was the time lapsed prior to the psychologist's last administration of the NART test

in practice ($t = 2.44$, $p<.016$), and the order with which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first) ($t = 2.97$, $p<.004$). The remaining factor considered to be a possible predictor of NART scoring errors made by psychologists had no significant effect. This factor was the number of times the psychologists had administered the NART test ($t = -1.38$, $p<.170$). One other factor was considered, the number of NART variation errors psychologists made at stage one. This was also found to be not significant ($t = -1.39$, $p<.168$).

In light of the regression analysis for Gp1 errors made by psychologists, three variables were chosen to be used as covariates in the analysis of psychology group differences in scoring NART errors for Gp1. The most significant variable was the number of NART errors made by psychologists at stage 1 of the study. The other two variables used as covariates were the time lapsed prior to the psychologist's last administration of the NART test in practice and the order with which the psychologists rated the 10 general population participants. Although the two variables were not significant at .001 they were significant at .05. This may be attributed to a Type I error (although unlikely due to the sample size). However, as a precautionary measure the variables were used as covariates.

The same regression analysis was carried out for the total number of errors made by psychologists scoring the second set of five subjects (Gp2) (post introduction of either learning aid). For this analysis the aid for scoring used by the psychologists was also entered as a variable. The only two significant variables to enter the equation at .001

were the psychologists NART error scores from stage 1 ($F(1, 106) = 122.5$, $p<.000001$) and the aid for scoring used by psychologists prior to rating the second set of five participants (Gp2) ($F(1, 106) = 19.0$, $df = 1$, $p<.00001$). When combined, the two variables accounted for 67% of variance (based on the adjusted R square statistic) in psychologist errors when scoring Gp2 for NART errors. A third variable, the order of rating the two sets of five participants, was just significant at the .05 level ($t = -2.02$, $p<.046$). The order variable and the more significant NART variable were used as covariates when analysing the psychologists' rating of the second set of five subjects (Gp2) and the effect of the introduction of the NART scoring aids.

Analysis of Variance

1. Analysis of effects on psychologists' NART error scoring pre and post introduction to scoring aid (audio taped NART or manual pronunciation guidelines) using identified covariates.

A one-way ANOVA was used to analyse the between group effects of psychologists by rating the total number of errors for the first set of five participants (Gp1 errors). Three covariates were included in the analysis namely; the number of NART errors made by psychologists at stage 1 of the analysis, the time lapsed prior to the psychologists' last administration of the NART test in practice and the order with which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first). The covariates were all significant in accounting for variation in error

scoring by psychologists on Gp1 errors (see table 7 for levels of significance). The difference in psychologist groups (ie. Clinical, Trainee and Assistant) only just reached minimal level of significance ($p<.0.04$).

A 3X2 ANCOVA was used to analyse the psychologist between group effect and the learning aid they used prior to their rating of the total number of errors for the second set of five participants (Gp2 errors). The main finding was that a significant interaction effect was evident between the aid for scoring and group differences (see table 7 for level of significance). The introduction of the scoring aids had a significant effect on reducing error scoring by psychologists. The difference in psychologist groups (ie Clinical, Trainee and Assistant) was not a significant factor on its own. Two covariates were included in the analysis namely; the number of NART errors made by psychologists at stage 1 of the analysis and the order that psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first). The covariates were all significant in accounting for variation in error scoring by psychologists on Gp2 errors.

2. Analysis of differences in psychologists' NART error scoring pre and post introduction to scoring aid (audio taped NART pronunciations or Manual Guideline pronunciations).

A one-way ANOVA was used to investigate the difference between psychologist groups in NART error scoring of the 10 general population participants. The psychologists were compared by two groups; Clinical, Trainee and Assistant

Table 7: Number of NART Scoring Errors made between the three Psychologist groups scoring the 10 general population participants pre- and post- exposure to the scoring aid (either tape or manual guideline)

	Total NART Scoring Errors 1 (pre) F (p)	Total NART Scoring Errors 2 (post) F (p)
Covariates		
NART Errors (made in stage 1)	129.4 (p<.0001*)	127.2 (p<.0001*)
Order of sets of 5 rated	10.7 (p<.001*)	5.02 (p<.027*)
Time since last administerd NART	5.54 (p<.020*)	Not entered
Combined Covariates	60.6 (p<.0001*)	45.5 (p<.0001*)
Main Variables		
Group Differences	3.23 (p<.043*)	0.05 (p<.953; NS)
Use of scoring/learning aid	Not entered	48.5 (p<.0001*)
Interaction effects (main variables)	-	4.90 (P<.009*)

* = **Significant** **NS** = **Not significant**

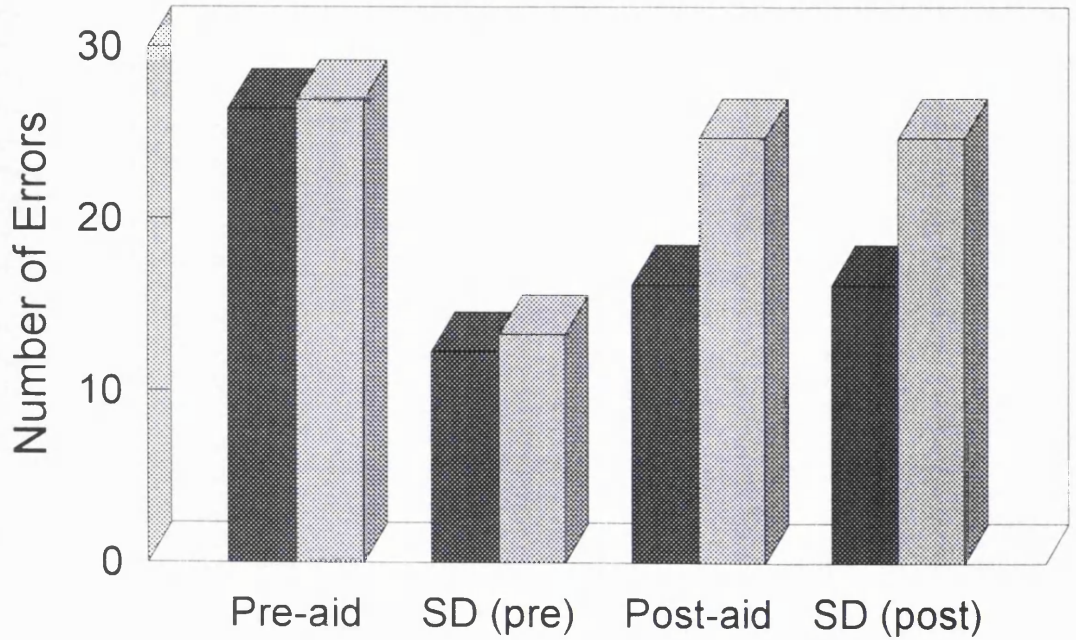
psychologists using the tape (N=54) and those using the NART manual guideline (N=54). The groups were compared for error scores pre- (NART scoring error 1) and post- (NART scoring errors 2) exposure to the scoring aids. There was no significant difference between the two groups when scoring the first set of five subjects ($F(1, 106) = .044$, $p<.834$). However, following the introduction of the scoring aids prior to psychologists rating the second set of participants a highly significant difference in NART scoring errors were found ($F(1, 106) = 19.0$, $p<.00001$). The psychologists using the tape as a scoring aid made significantly less scoring errors compared to the psychologists using the NART manual guideline. The mean difference in errors between the two groups equalled -8.5 (see chart 6 for means and SD).

The other main significant finding was the difference in standard deviation (SD) in NART scoring errors made by psychologists pre and post exposure to either scoring aid (see chart 6). The SD for the NART scoring errors made by psychologists using the tape does not vary significantly pre- and post-exposure to the tape. However, the SD for the scoring errors made by psychologists using the NART manual guideline significantly increases. The introduction of the manual guideline caused a wider distribution of error scores.

The second analysis of difference in psychologists' NART error scoring compared each group by type of psychologist (Clinical, Trainee and Assistant), the type of aid used (tape or manual guideline) and the reduction of NART scoring errors made. A within-subjects design was used comparing each psychologists pre- and post-use of aid rating

Chart 6: Psychologists Mean Error

score rating of subject groups pre & post exposure to aids



Tape

Manual Guideline

N=108

SD = Standard Deviation

of errors. A t-test was used to check for any differences in scoring. It was clearly evident from the analysis that the psychologists using the tape made significantly less errors after exposure to the tape. The three groups of psychologists using the manual guideline showed no significant difference in their level of NART error scoring following exposure to the NART manual guideline (see table 8 and chart 7 for levels of significance and means).

Psychologists' aid preference for scoring and learning NART word pronunciation

After the psychologists had rated the 10 general population participants they were asked which aid for scoring NART errors and learning word pronunciation would be most useful in practice. The two options offered were the NART pronunciation manual guideline or the audio taped NART pronunciation guideline. 97% of participating psychologists (N=101) supported the use of the audio taped version for use in learning and scoring the NART words. 6% of psychologists (N=7) supported the use of an audio-taped word pronunciation guide combined with a NART manual guideline type pronunciation guide.

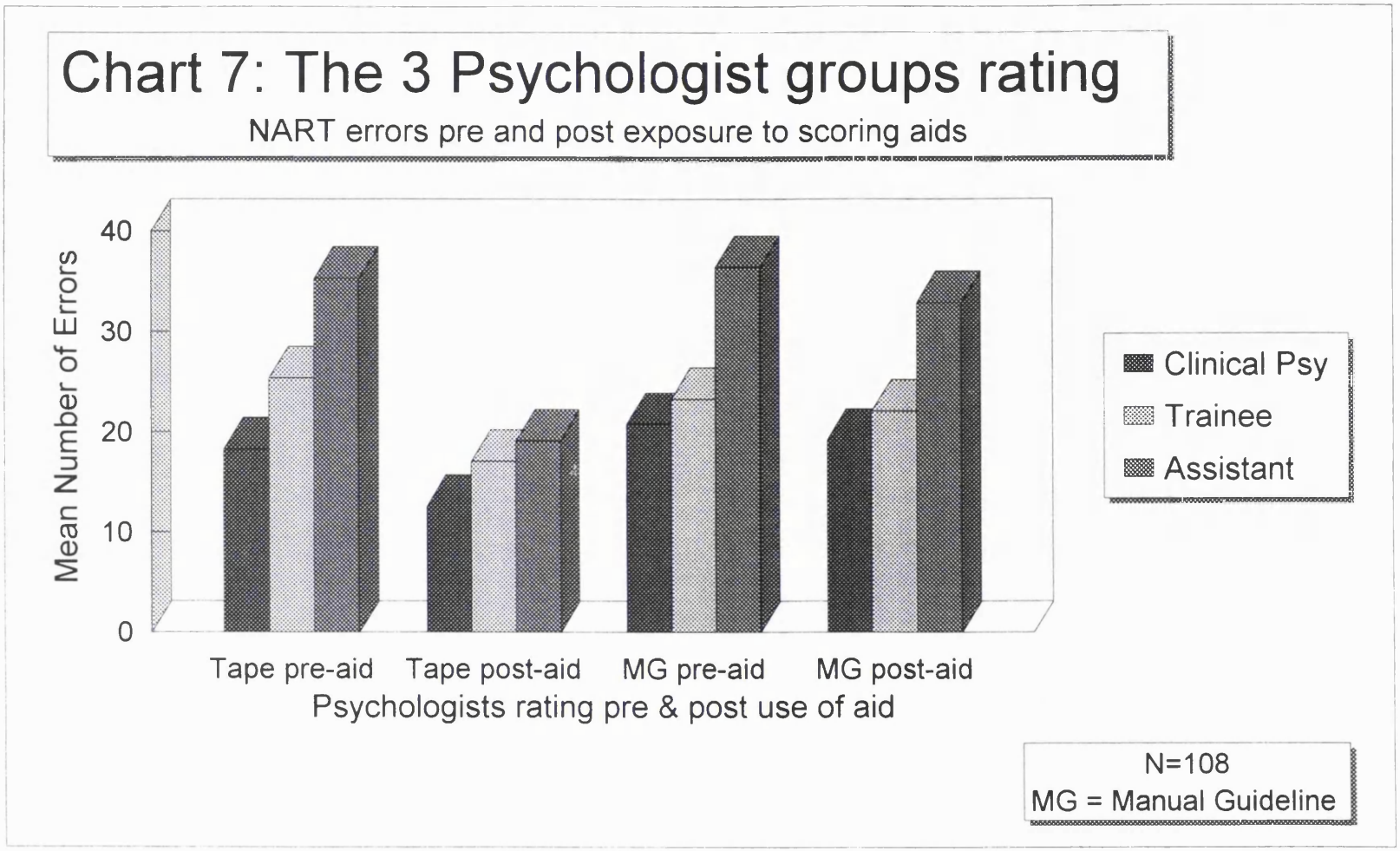


Chart 7: The 3 Psychologist groups rating
NART errors pre and post exposure to scoring aids
Mean Number of Errors
40
30
20
10
0
Tape pre-aid
Tape post-aid
MG pre-aid
MG post-aid
Psychologists rating pre & post use of aid
Clinical Psy
Trainee
Assistant
N=108
MG = Manual Guideline

Table 8: Number of NART Scoring Errors made within each of the three Psychologist groups scoring the 10 general population participants pre- and post-exposure to the scoring aid (using either the tape or manual guideline)

	Total NART Scoring Errors 1 (pre)	Total NART Scoring Errors 2 (post)	Level of Significance t (p)
Clinical Psychologists (N=36)			
Tape (N=18)	18.3 (7.9)	12.6 (2.9)	3.06 (.007*)
Manual Guideline (N=18)	20.8 (8.1)	19.2 (7.6)	1.34 (.196; NS)
Trainee Psychologists (N=36)			
Tape (N=18)	25.4 (10.7)	17.1 (9.4)	7.53 (.000*)
Manual Guideline (N=18)	23.3 (9.2)	22.1 (9.7)	0.76 (.458; NS)
Assistant Psychologists (N=36)			
Tape (N=18)	35.3 (11.9)	19.1 (9.8)	8.28 (.000*)
Manual Guideline (N=18)	36.4 (15.9)	32.9 (12.7)	1.79 (.092; NS)

*** = Significant NS = Not significant**

Discussion

The study findings are numerous and have implications for the use of the NART in both research and clinical practice. The first hypothesis postulated was that all three groups of psychologists would make errors in NART word pronunciation. This hypothesis was confirmed in stage 1 of the study when each of the participating psychologists were administered the NART and rated for errors (out of 50) and NART word variation errors (out of 57). Only 4 psychologists out of 108 participants (3.7%) were able to pronounce the standard 50 NART words without making any errors. Furthermore, no psychologist was able to produce the full 57 word variations presented in the NART manual pronunciation guideline. In fact, no psychologist scored below 5 errors on the full 57 word variations.

The significance of these findings is associated with the method used to administer the NART test in practice. The method generally used for the administration and scoring of the NART test provides a great deal of scope for making errors in scoring. The assessor would normally administer the test by sitting with the subject and asking the individual to read the words aloud. The assessor would then either rate each word as the person read or make a note of the pronunciation for later scoring (mainly if they are not sure of the pronunciation). Using this method of rating, the assessor would need to feel very comfortable with NART word pronunciation and variations to rate the patient/study participant. This difficulty is compounded when attempts are made to deal with problems associated

with other influences (eg. accent variations). Furthermore, there is an assumption that psychologists administering the NART would know the correct pronunciation of the words used in the NART test. This assumption has no basis in fact. It is evident from this study that many psychologists are not be able to pronounce all the NART words correctly (and with variations) and consequently make a wide range of errors when scoring the NART.

It may be argued that some psychologists often use the NART manual pronunciation guideline in the test situation but because the study design prevented its use, the situation was not reflective of the actual test situation. However, perceived positive benefits of the presence of the NART manual pronunciation guideline pre-supposes that the manual guideline is of benefit to psychologists in understanding the correct word pronunciation. The study findings indicate that exposure to the manual pronunciation guideline produced no significant reduction in errors made by psychologists when rating study participants or when pronouncing the words themselves. The participating psychologists were actively encouraged to spend as much time as they needed reviewing the manual guideline prior to participation in the study. Furthermore, the NART manual guideline is generally used by psychologists in the test situation to check words they are unsure about. However, the pronunciation of words that psychologists feel unsure about may be wrong in some cases. Consequently, the psychologist would not seek any verification of those words and would continue to make errors in scoring even with the assistance of the manual pronunciation guideline.

The type of cognitive processing used by psychologists when scoring NART errors and accounting for variations in word pronunciation can be described using *prototype* (or characteristic-attribute) theories. The prototype view is that categories are organised around central prototypes (Eysenck & Keane, 1990). The *exemplar* prototype represents the prototype in terms of the best member (or small set of best members) of the concept (Brooks, 1978; Medin & Shaffer, 1978; Hintzman & Ludlum, 1980). For example, the prototype of a NART word category for DEMESNE might be the pronunciation *di-man* exemplar or a set of exemplars (*di-man, di-men).* Here a word is a member of a category to the extent that it is close to the best examples of the concept or in the case of NART scoring matching the word category (ie word and/or word variations).

The prototype theory accounts for the fuzzy areas of NART scoring with psychologists scoring NART words on the basis of it sounding similar to the word (ie. close to the main word exemplar prototype) so the pronunciation was given a correct score even though it did not perfectly match the exemplar prototype word. The problem psychologists experienced with word variations would appear to be a limited set of word prototypes (ie. the psychologists hold only a single word pronunciation prototype without any variation on the main word prototype). Consequently, if the pronunciation provided by the patient/study participant does not match the single word prototype then it is scored incorrectly. The problems participating psychologists experienced in recalling the word pronunciations and variations would appear to be more associated with the interpretation of word pronunciation using the NART manual

word pronunciation guide. It is unlikely that all the participating psychologists were experiencing cognitive problems with their *auditory analysis system* (used to extract phonemes or other sounds from the speech wave), *auditory input lexicon* (contains information about spoken words known to the listener) or their *speech output lexicon* (which stores spoken forms of words) all part of the cognitive system which processes spoken words (Ellis & Young, 1988) However, if the spoken form of a word (or word variation) given by the patient/study participant is not stored in the auditory input lexicon (ie. possibly due to the inability to interpret the NART manual pronunciation guideline) then the word pronunciation would be unfamiliar to them and consequently scored incorrect.

The problems psychologists experienced in providing the full range of acceptable word pronunciations (presented in the NART manual pronunciation guideline) may lead to scoring correct word variations incorrect. If psychologists are either not aware of, or only able to hold one possible variation of the word in the test situation, then any variations of the *prototype* word held by the psychologist would not be matched with their own interpretation of the correct pronunciation. Consequently, the psychologist would score the patient/study participant wrong when the person may have given one of the correct variations on the word. This problem is further compounded when psychologists variations on words are incorrect variants. It is clear from the study that participating psychologists gave incorrect variations on words which they perceived as correct. Consequently, if a patient/study participant gave a correct variation which

again did not match the psychologist's prototype word(s) then the psychologist would score them wrong when in actual fact a correct response was given.

The implications of the prototypes which psychologists hold for words in terms of correct/incorrect pronunciation and the lack of significant reduction in errors made by psychologists following exposure to the NART manual guideline creates an opening for a wide range of possible errors when scoring individuals using the NART. The mean NART errors made by psychologists (out of 50) and the mean NART variation errors (out of 57) at stage 1 of the study provided a combined mean for the psychology groups of 8 and 14 errors respectively. These mean error predictions were made without considering any other influences/biases that may affect a psychologist when scoring NART errors.

Participating psychologists reported a number of influences affecting their rating of NART errors based on the scoring of the 10 general population participants in the study and from their general practice experience. Some of these factors were found to be significantly associated with the errors participating psychologists made when scoring the first five general population study participants pre-exposure to any scoring aids. The factors include: the number of NART errors psychologists made at stage 1 of the study when completing the NART test themselves, the number of times the psychologists had administered the NART test, the time lapsed prior to psychologists last administration of the NART test in practice and the order in which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first).

Not surprisingly, the most significant factor associated with the total number of NART scoring errors made by psychologists is the number of NART errors they made themselves. This single factor accounted for 68% of variance in scoring errors made by psychologists. The association between the NART pronunciation errors and NART scoring errors made by psychologists makes sense. If a psychologist makes errors on the NART test themselves then those errors will more than likely affect their subsequent scoring of patients/study participants using the NART. This again raises the importance of the provision of adequate aids for learning and scoring NART word pronunciation. A possible argument that may be presented is that the participating psychologists were ill-prepared for the use of the NART and were not using the manual pronunciation guideline adequately. However, in 87% of cases the participating psychologists had used the manual guideline as a guide to learning the correct word pronunciation. It may be possible that psychologists do not use the guideline frequently enough to refresh their understanding of word pronunciation. However, this then again raises the issue of how much benefit psychologists gain from using the manual guideline. These study findings indicate that little benefit is gained from using the NART manual pronunciation guideline.

The number of times a psychologist had administered the test prior to involvement in the study was also significantly inversely associated with NART errors and NART scoring errors made by psychologists. Although this finding was predicted pre-data collection it is still a slightly surprising finding. One might assume that if a person is making errors on the NART test from the beginning of test usage then the on-going

use of the test may just compound the errors. However, an alternative explanation that may account for the study finding is that increased frequency of NART test usage may be associated with the interest psychologists would take in familiarising themselves with the NART word pronunciation by following up queries about word pronunciation and through word practice. The frequent use of the test also exposes them to a wider range of pronunciations and accents (etc).

The other significant influences on NART errors/scoring errors made by psychologists is the time lapsed since the NART test was last administered prior to involvement in the study and the order in which the psychologists rated the 10 general population participants (ie 1-5 first or 6-10 first). A significant positive association between the recency of test administration and reduced NART errors is not particularly surprising. The more recent the exposure to the NART test experienced by psychologists the increased likelihood of familiarity with the NART words and their pronunciation. However, the recency and frequency of test use are correlated and it is likely that one variable subsumes the other.

An order effect was anticipated pre-data collection and order of presentation for rating of the two sets of five participants (A and B) were reversed in 50% of cases in an attempt to control the order effects. Participating psychologists made less errors when scoring group A compared to group B. This was a consistent trend found when scoring NART errors pre- and post-introduction of the scoring aids (tape or manual guideline). Due to the consistency of reversing the order (in 50% of cases) this should

have controlled for a significant amount of the order effect. The order effect was used as a covariate when using Analysis of Variance statistics to identify differences in NART error scoring between the psychologist groups. A within-subjects design may have been implemented using one set of five general population participants to be rated pre- and post-exposure to the scoring aids. However, this method would have resulted in the need to control for a practice effect. Whichever method is used, some control for study design effects needs to be implemented.

One of the most interesting findings from the NART error ratings was the list of factors associated with how psychologists scored word pronunciation right or wrong. The list consisted of adjustments for accents, whether a person was confident/hesitant in their word pronunciation, combination, previous rating, learning, positive and class biases. These factors were reported in the study findings to have no significant effect on psychologist scoring of NART errors. However, this does not mean that such factors had no effect on scoring. The nature of this study design does lend itself to close scrutiny of these factors. The way in which the factors were grouped probably diluted the prevalence of the two key factors; adjustment for accents and rating NART errors on the basis of the person's confidence/hesitance on word pronunciation. The combination of the biases group consisted of mainly accent and confidence associated influences (combined with other influences). If these had been allocated to the main accents and confidence groups it would have increased their numbers significantly. However, because the influences were reported with other influences by the same psychologist it was difficult to allocate them to the either one of the main groups.

These factors definitely warrant closer investigation.

The identification of various influences and biases reported by psychologists when using the NART test has implications for the use of the test in practice. The method of administering and scoring the test provides a great deal of scope for allowing biases to influence scoring. The psychologist is placed in an evaluative situation whereby they make a judgement about whether the patient/study participant made a correct/incorrect word pronunciation. The amount of information the psychologist would be registering about the person verbally/non-verbally, consciously/unconsciously is most likely to effect their error scoring in some way. However, these influences would not be applicable to other proposed estimates of premorbid ability. The Spot-The-Word test (STW; Baddeley et al, 1993) would not be affected in any way by these types of influences. The STW makes no evaluative demands on the psychologist. The person taking the test completes the procedure independently of the psychologist's interpretation of whether a list of paired words are real or non-words. It is the person taking the test who is making the judgement not the person administering the test. Consequently, in controlling for these types of biases/influences the STW test is far superior than the NART test. However, the STW test currently lacks the recognised validity of the NART in estimating premorbid ability in various clinical groups.

The most significant finding of the study in terms of implications for learning how to pronounce and score the NART words is the effect of the NART manual and audio

taped pronunciation guidelines on participating psychologists' ability to score general study participants for NART errors.

There was no significant difference in NART error scoring made by the two groups of participating psychologists (using the tape or manual guideline) pre-exposure to either scoring aid. This is an important finding because it shows that both groups of psychologists were making a similar number of errors at this stage prior to any influence from the scoring aids. The factors influencing psychologists at this stage are their previous methods of learning word pronunciation prior to involvement in the study (NART manual guideline, instruction from supervisor etc.)

When comparing the NART scoring errors made by the two groups of psychologists rating the second set of tapes (post-exposure to the scoring aids) the study findings were unequivocal. The audio taped pronunciation guideline significantly reduced the number of NART scoring errors made by participating psychologists. In comparison the NART manual guideline produced no significant change in error scoring. The findings were consistent when the participating psychologists were divided into three groups (Clinical, Trainee and Assistant psychologists). In each group, the psychologists using the taped pronunciation guideline performed better than the psychologists using the NART manual guideline in terms of scoring errors made. These findings raise the question of whether the current NART manual guideline is an adequate aid for psychologists to use in practice.

At this stage it would be useful to scrutinise the NART manual pronunciation guideline in more detail raising some of the problems associated with its current format. The manual guideline was developed for use in the initial standardisation sample (Nelson, 1982). The pronunciation guideline has been a feature of the NART manual throughout its early development and up to the restandardisation of the NART (Nelson & Willison, 1991) and its present day use. The pronunciation guideline is a very important part of the NART test. It provides the actual criteria for psychologists to follow regarding what is acceptable or unacceptable as a correct pronunciation. Consequently, all research and normative data using the NART are based on such criteria. If there are problems with the pronunciation guide then there will be associated problems with the NART test in general. This has significant implications for both previous and current research and clinical practice. The study findings have shown that the current NART manual pronunciation guideline is not adequate for use in practice, as it does not significantly reduce psychologists NART scoring errors. This finding has certain implications as it throws doubt on both previous and current use of the NART both for research and clinical practice.

There are further problems with the NART manual pronunciation guideline. The guideline was initially adapted from the Penguin Dictionary (Nelson & O'Connell, 1978) and then the Chambers Dictionary for the current restandardised version of the NART (Nelson, 1982 pg. 4; Nelson & Willison, 1991). For example, in the Penguin Dictionary three alternative pronunciations for the word *Campanile* were considered acceptable (*Kam-pan-e'le, Kam-pan-e'la, and Kam-pan-il*). However, in the current

test only the first two pronunciations are acceptable (Bucks et al, 1996). Furthermore, in the English Pronouncing Dictionary (a phonetics dictionary; Jones, 1997; originally written by Susan Ramsaran, 1991) the first pronunciation is considered acceptable. Therefore in three different dictionaries there are three variations in interpretation of what is considered a correct pronunciation. This point is indicative of the precarious nature of defining correct word pronunciation. If at one point (or one dictionary) three variations of a word pronunciation are correct and at another point (or another dictionary) one word variation is acceptable, this provides a wide scope for error in deciding what is considered correct and incorrect word pronunciation. The inconsistency between dictionaries on word pronunciation provides problems for the development of any pronunciation guideline for scoring each individual's performance on word reading. The question at this point is on what basis was the Chambers Dictionary chosen to provide the criteria for the NART manual pronunciation guideline? Three renowned alternative dictionaries (The Concise Oxford Dictionary, 9th ed., 1995; the English Pronouncing Dictionary, Jones, 1997; and the Longman's Pronunciation Dictionary, Wells, 1997) one standard and two phonetics dictionaries all indicate that *Campanile* has only one word pronunciation variation (*Kam-pan-e'le*). Therefore, the consensus between dictionaries would seem to indicate that the one word pronunciation variation is more valid than the two word variations presented in the NART manual pronunciation guideline.

The development of any future guideline would need careful consideration about where the criteria for correct pronunciation should be taken from, and who decides

what is the most suitable source of information to provide the criteria. It would seem sensible to seek guidance in dealing with these points possibly through consultation with specialists in English Language in deciding what is the most suitable source criteria for a word pronunciation guideline.

The problems with the NART manual pronunciation guideline and the apparent difficulty psychologists seem to experience when attempting to learn the correct NART word pronunciations (and variations) raise two questions: why do psychologists experience problems learning the NART words? and what alternative approaches may be used to facilitate NART word learning? In attempting to shed some light on these questions it is useful to review some of the educational tensions associated with the test.

The NART test can make people feel very uncomfortable. The participating psychologists often openly expressed their feelings about the threatening nature of the study design mainly associated with the stage 1 procedure. The psychologists were requested to read the NART words (with variations) into an audio-tape at this stage. The *anxiety factor* associated with tests of intelligence and particularly core skills like word reading/pronunciation is compounded when highly professional groups are asked to take part in a study evaluating their own abilities. It was evident from the study that the qualified clinical psychologists were more apprehensive about participating in the study compared with trainee and assistant psychologists. This is reflected in the number and type of psychologists that refused to participate in the study.

The fear factor may be one of the problems associated with the errors in NART pronunciation and subsequent NART scoring errors. Seeking understanding about the pronunciation of words may be perceived by some people as a negative reflection of their educational and professional standing. Therefore, the protection of that status may prevent psychologists seeking instruction in the correct pronunciation of the NART words.

Another possible explanation as to why psychologists may experience problems learning the NART words is associated with their inability to understand phonetics and phonetic symbols. It is assumed that psychologists have these skills. However, the fact that psychologists responded more positively (in terms of making less scoring errors) to a taped version of the NART words compared to using the NART manual guideline, suggests that the word pronunciation was learnt more easily when using an audio-taped version of the words.

The audio taped version of the NART words provides a less threatening method of learning the word pronunciations. The psychologist is able to listen to how the word sounds making it easier to recognise the correct sound of the word in practice. The taped version of NART word pronunciations also removes the assumption that psychologists have some expertise in phonetics and in the translation of phonetic coding. The primary concern for the psychologist when using the NART test is to be able to score the word pronunciations with some confidence and therefore obtain a

more valid score estimate. It is not in any person's interest to make the road to feeling confident with word pronunciation a difficult one. If the guideline can be improved then this option should be considered to make the test more user friendly. When each psychologist was presented with both the audio-taped version of the correct word pronunciation and the NART manual pronunciation guideline all of the participating psychologists indicated that they would benefit most from using the audio-tape (either on its own or in combination with the manual guideline) as an aid to scoring/learning the NART words.

The NART test could be described as a *victim of its own simplicity.* The test appears to be a simple, quick to administer, test. The preparation for its administration seemingly does not demand the amount of time allocated to more complex tests like the Wechsler scales. However, the complexities of the test are found within the understanding of the various word pronunciations, dealing with accents and other influences and the method of test administration and scoring. Due to the test's apparent simplicity it perhaps does not receive the preparation time it deserves.

Clinically reported NART and Wechsler scale intelligence estimates can have serious implications for the care planning process and the subsequent provision of services. This is most evident when the test is used with the WAIS-R as a predictor of intellectual decline following the onset of illness/or injury. The time allocated by psychologists to familiarise themselves with word pronunciation prior to administering the test is minimal in comparison to other tests. However, for such a short, simple

test, it must be one of the most frequently discussed measures in both clinical psychology circles and other associated professional groups. It was surprising how many participating psychologists stated *'I was discussing the NART and the word pronunciations with some colleagues recently....'*

The debate continues about whether a test can estimate premorbid ability. Comments about the NART made by participating psychologists were varied and were generally associated with its usefulness, reliability and validity. Some psychologists value the test highly whereas others criticise the test by pointing out its weaknesses (educationally, culturally, and whether it or any other test can actually estimate premorbid intelligence). One might argue that the study findings are not reflective of the well quoted high inter-rater, split-half and test-retest reliability figures (Nelson, 1982; O'Carroll, 1987; Crawford et al., 1989a; Schlosser & Ivison, 1989) presented in various studies (ie. if psychologists are making so many NART scoring errors, then why are the NART reliability figures so high?).

This point can be explained by highlighting the nature of reliability tests. The reliability tests only identify agreement between psychologists (or other participant groups) on the rating of an individual's score either between raters or by the same rater scoring a person more than once following a delay. If there is consistency in scoring between the raters then a high positive correlation will be achieved. However, a consistency in scoring does not guarantee that the raters are scoring correctly or incorrectly it only provides information about consistency in agreement. Consequently, psychologists

could (and are) making errors in NART scoring and may still produce high reliability figures if the psychologists are making consistent errors in scoring. The point was made earlier that less reliable words (ie. those causing most disagreement between raters) were removed from the test both in the standardisation of the NART (Nelson, 1982) and the development of revised versions of the NART (NART-R; Crawford, 1990). Although this procedure may help to ensure some agreement between raters, it may also compound the scoring error problem. This is because the words remaining are those which raters may be consistently scoring right or wrong. Perhaps then a more useful approach to investigating future reliability tests on the NART should consider not only consistent agreement between raters but also whether the agreement between raters are correct or incorrect.

The wider implications of the present findings is that the current NART manual pronunciation guideline is not suitable in its current form. If the guideline is to be replaced then a complete restandardisation of the NART test would need to be completed. This is due to the fact that the standardisation sample and all existing normative data are based on the current guideline. This has implications not only for the original UK version of the NART test but also variations of the test using the same format (the NART-R, Crawford, 1990; the North American Adult Reading Test - Blair & Spreen, 1989; and the Dutch Adult Reading Test; Schmand et al, 1991) and the same words (the Cambridge Contextual Reading Test; Beardsall & Huppert, 1994).

A number of improvements could be made to the test by removing the words in the test that have a number of word variations and replacing them with single variation words. This would reduce the likelihood of errors associated with word variations. If the psychologist had only one variation on each word to learn it would more than likely increase the consistency/reliability in scoring a word pronunciation right or wrong.

The NART would also benefit from a closer investigation into the development of the NART pronunciation guideline possibly providing both an audio-taped version of the pronunciation guideline and a text style manual guideline. This would lead to a reduction in NART scoring errors made by psychologists based on the current study findings.

Another possible improvement relates to the problems psychologists encounter when dealing with accents. The study findings indicate that in a number of cases each psychologist is making their own idiosyncratic adjustment for accents when scoring NART errors. This will have implications for standardised scoring and the development of normative data. If there is a need to make adjustment in scoring NART errors for people with accent variations then a standardised approach needs to be identified and implemented. The problem with accents may be alleviated by the development of a pronunciation guideline for each region. However, this proposal would be impractical. A more viable proposition would be the identification and use of words that are shown to be less affected by regional dialects. Consequently, if a

standardised approach to dealing with accent variations were adopted then psychologists would less inclined to make their own idiosyncratic adjustments. The recommendations made would form the basis for future research projects.

Conclusions

The main finding of the study is that there was no significant difference in NART scoring errors made between psychologist groups' (using an audio taped version of the NART words or the NART manual pronunciation guideline when scoring) pre-exposure to either scoring aid. However, a highly significant difference in NART scoring errors was found post-exposure to the scoring aids. The psychologists using the tape as a scoring aid made significantly less scoring errors compared to the psychologists using the NART manual guideline. The standard deviation for the NART scoring errors made by psychologists using the tape does not vary significantly pre and post exposure to the tape. However, the standard deviation for the scoring errors made by psychologists using the NART manual guideline significantly increases. The introduction of the manual guideline caused a wider distribution of error scores. An interaction effect was found between the three psychologist groups and the scoring aid used. The findings throw some doubt on the current and past use of the NART both for research and clinical practice and the use of the NART manual pronunciation guideline as criteria for scoring word pronunciation.

References

Austin, M-P., Ross, M., Murray, C., O'Carroll, R.E., Ebmeier, K.P., & Goodwin, G.M. (1992). Cognitive function in major depression. *Journal of Affective Disorders*, **25**, 21-30.

Baddeley, A.D., Emslie, H. & Nimmo-Smith, I. (1990). *The Speed and Capacity of Language Processing (SCOLP) Test*. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Baddeley, A.D., 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**, 55-65.

Barker, C., Pistrang, N. & Elliot, R. (1994). *Research Methods in Clinical and Counselling Psychology*. Chichester, West Sussex: Wiley & Sons.

Beardsall, L. & Brayne, C. (1990). Estimation of premorbid intelligence in an elderly community: A prediction analysis using a shortened NART. *British Journal of Clinical Psychology*, **29**, 83-90.

Beardsall, L. & Huppert, F.A., (1994). Improvement in NART word reading in demented and normal older persons using the Cambridge Contextual Reading Test. *Journal of Clinical and Experimental Neuropsychology*, **16(2)**, 232-242.

Blackmore, L., Crawford, J.R., & Simpson, S.A. (1994). Current and premorbid intelligence in Huntington's disease. *Proceeds of the British Psychological Society*, **2,** 57.

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

Bolter, G., Gouvier, W., Veneklasen, J. & Long, C.J. (1982). Using demographic information to predict IQ: A test of clinical validity with head trauma patients. *Clinical Neuropsychology*, **4**, 171-174.

Broek, van den, M.D. & Bradshaw, C.M. (1994). Detection of acquired deficits in general intelligence using the National Adult Reading Test and Raven's Standard Progressive Matrices. *British Journal of Clinical Psychology*, **33 (4)**, 509-515.

Brooks, L. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B.B. Lloyd (Eds), *Cognition and Categorisation*. Hillsdale, N.J.: LEA.

Bucks, R.S., Scott, M.I., Pearsall, T., & Ashworth, D.L. (1996). The short NART: utility in a memory disorders clinic. *British Journal of Clinical Psychology*, **35 (1)**, 133-141.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37-46.

Coolidge, F.L., Peters, B.M., Brown, R.E., Harsch, T.L., & Crookes, T.G. (1985). Validation of a WAIS algorithm for the early onset of dementia. *Psychological Reports*, **57**, 1299-1302.

Crawford, J.R., Besson, J.A.O., Parker, D.M., Sutherland, K.M., & Keen, P.L. (1987). Estimation of premorbid intellectual status in depression. *British Journal of Clinical Psychology*, **26**, 313-314

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

Crawford, J.R., Stewart, L.E., Calder, S., Ebmeier, K., Mutch, W., & Besson, J.A.O. (1988b). *Estimation of premorbid intelligence in idiopathic Parkinson's disease.* Israel, book of abstracts, 9.

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

Crawford, J.R. (1990). *The estimation of Premorbid Intelligence.* PhD Thesis, University of Aberdeen.

Crawford, J.R., Nelson, H.E., Blackmore, L., & Cochrane, R.H.B. (1990a). Estimating premorbid intelligence by combining the NART and demographic variables:

An examination of the NART standardisation sample and supplementary equations. *Personality and Individual Differences*, **10**, 585-587

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

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

Crawford, J.R., Allan, K.M., Jack, A.M., Morrison, F.M., & Parker, D.M. (1991). The short NART: cross-validation, relationship to IQ and some practical considerations. *British Journal of Clinical Psychology*, **30**, 223-229.

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

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

Cummings, J.L., Houlihan, J.P., & Hill, M.A. (1986). The pattern of reading deterioration in dementia of the Alzheimer's type: observations and implications. *Brain & Language*, **29**, 315-323.

Dunkley, G. & Rogers, D. (1994). The cognitive impairment of severe psychiatric illness: A clinical study. Reported in: *The Neuropsychology of Schizophrenia.* Eds; David A.S., & Cutting J.C., LEA, Hove, UK (1994). Chapter 8, pg 131-144.

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

Egan, V.G., Crawford, J.R., Brettle, R.P., & Goodwin, G.M (1990). The Edinburgh cohort of HIV-positive drug users: Current intellectual functioning is impaired, but not due to early AIDS Dementia complex. *AIDS*, **4**, 651-656.

Ellis, A.W. & Young, A.W. (1988). *Human Cognitive Neuropsychology.* London: LEA.

Eysenck, M.W. & Keane, M.T. (1990). *Cognitive Psychology: A students handbook.* LEA: Hove & London.

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

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

Hintzman, D.L. & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory Cognition*, **8,** 378-382.

Hodges, J.R., Salmon, D.P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's Disease: Failure of access or degraded knowledge? *Neuropsychologia*, **30**, 301-314.

Howell, D.C. (1992). *Statistical methods for psychology (3rd Ed.)*. Belmont, California: Duxbury Press.

Jones, D. (1997). *The English Pronouncing Dictionary*. Cambridge: Cambridge University Press.

Jones, P., Guth, C., Lewis, S, & Murray, R. (1994). Low intelligence and poor educational achievement precede early onset schizophrenic psychosis. Reported in:

The Neuropsychology of Schizophrenia. Eds; David A.S., & Cutting J.C., LEA, Hove, UK (1994). Chapter 8, pg 131-144.

Karzmak, P., Heaton, R.K., Grant, I., & Matthews, C.G. (1985). Use of demographic variables to predict Full Scale IQ: A replication and extension. *Journal of Clinical and Experimental Neuropsychology*, 7, 412-420.

Klesges, R.C., Fisher, L., Vasey, M., & Pheley, A. (1985). Predicting adult premorbid functioning levels: Another look. *International Journal of Clinical Neuropsychology*, 7, 1-3.

Lees, A.J. & Smith, E. (1983). Cognitive deficits in the early stages of Parkinson's disease. *Brain*, **106**, 257-270.

Levene, H. (1960). Robust tests for the equality of variance. In I. Olkin (Ed), *Contributions to probability and statistics*. Palo Alto, CA: Stanford University Press.

Lezak, M.D. (1983). *Neuropsychological Assessment* (eds.). Oxford: Oxford University Press.

Matarazzo, J.D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.

Medin, D.L. & Shaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, **85,** 207-238.

Mockler, D.M., Riordan, J.M., & Sharma, T. (1996). A comparison of the NART (Restandardised) and the NART (Revised). *British Journal of Clinical Psychology*, **35 (4)**, 567-572.

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

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

Nelson, H.E. & O'Connell, A. (1978) Dementia: the estimation of premorbid intelligence levels using the new adult reading test. *Cortex*, **14**, 234-244.

Nelson H E, (1982). *National Adult Reading Test (NART): Test Manual.* Windsor: NFER-NELSON.

Nelson H.E., & Willison J.R. (1991). *National Adult Reading Test (part II): Test Manual.* Windsor: NFER-NELSON

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

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

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

O'Carroll, R.E., (1992). Predicting premorbid intellectual ability in dementia. *The Clinical Neuropsychologist*, **6,** 113-115.

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

O'Carroll, R.E., Walker, M., Dunan, J., Murray, C., A.P.R., Blackwood, D., Ebmeier, K.P., & Goodwin, G.M. (1992b). Selecting controls for schizophrenia research studies: The use of the National Adult Reading Test as a measure of premorbid ability. *Schizophrenia Research*, **8**, 137-141.

O'Carroll, R.E., Curran, S.M., Ross, M., Murray, C., Riddle, W., Moffoot, A.P.R., Ebmeier, K.P., & Goodwin, G.M. (1994). The differentiation of major depression

from dementia of the Alzheimer type using within subject discrepancy analysis. *British Journal of Clinical Psychology*, **33**, 23-32.

O'Carroll, R.E., Prentice, N., Murray, C., Beck, van, M.K., Ebmeier, K.P., & Goodwin, G.M. (1995). Further evidence that reading ability is not preserved in Alzheimer's disease. *British Journal of Psychiatry*, **167 (5)**, 659-662.

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

Paque, L. & Warrington, E.K. (1995). A longitudinal study of reading ability in patients suffering from dementia. *Journal of the International Neuropsychological Society*, **1(6)**, 517-524.

Patterson, K., Graham, N., & Hodges, J.R. (1994). Reading in Dementia of the Alzheimer type: A preserved ability? *Neuropsychology*, **9**, 395-412.

Ramsaran, S. (1991). *The English Pronouncing Dictionary (first edition)*. Cambridge: Cambridge University Press.

Riordan J.M., Mockler D.M., Sharma T., & Davies, N. (1994). The relationship between premorbid and current intelligence with clinical symptomatology in

schizophrenia. Presented at the winter workshop on schizophrenic research, Geneva, Les Diablerets. *Schizophrenia Research*, **11(2)**, 185.

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

Schmand, B., Bakker, D., Saan, R., & Louman, J. (1991). The Dutch reading test for Adults: a measure of premorbid intelligence. *Tijdschr Gerontol Geriatr*, **22(1)**, 15-19.

Sharpe, K.L. (1990). *Estimating premorbid intelligence using the National Adult Reading Test in Newfoundland.* MSc Thesis, Memorial University of Newfoundland.

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

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

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

Stebbins, G.T., Gilley, D.W., Wilson, R.S., Bernard, B.A., & Fox, J.H. (1990b). effects of language disturbances on premorbid estimates of IQ in mild dementia. *Clinical Neuropsychologist*, **4**, 64-68.

Struben, E.A.M., & Tredoux, C.G. (1989). *The estimation of premorbid intelligence: The National Adult Reading Test in South Africa.* Paper presented at the fourth National Congress of the Brain and Behaviour Society, Durban, South Africa.

The Chambers Dictionary (eds). Edinburgh: W & R Chambers Ltd.

The Concise Oxford Dictionary (9th edition) (1995). Oxford: Oxford University Press.

Wechsler D, (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

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

Wells, J.C. (1997). *Longman's Pronunciation Dictionary (eds)*. Harlow, Essex: Longman.

Wiens, A.N., Bryan, J.E., & Crossen, J.R. (1993). Estimating WAIS-R FSIQ from the National Adult Reading Test-Revised in normal subjects. *The Clinical Neuropsychologist*, 7, 70-84.

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

Wilson, R.S., Rosenbaum, G. & Brown, G. (1979). The problem with premorbid intelligence in neuropsychological assessment. *Journal of Clinical Neuropsychology*, **1**, 49-53.

Appendix 1

CONSENT FORM

TITLE OF PROJECT: ***Psychologists and their administration and scoring of the NATIONAL ADULT READING TEST (NART)***

The participant should complete the whole of this form him/herself

	Please Delete as Necessary
Was information about the study, and your role in it clearly explained to you?	YES / NO
Have you had the opportunity to ask questions and discuss the study?	YES / NO
Have you received satisfactory answers to your questions?	YES / NO
Have you received enough information about the study?	YES / NO

Who have you spoken to? Dr / Mr / Ms / Mrs _____

Do you understand that you are free to withdraw from the study

1. At any time
2. Without having to give a reason YES / NO

Do you agree to take part in this study?

Signed (signature) _____ Date _____

NAME (IN BLOCK LETTERS) _____

Appendix 2a: Scoring Template for subjects 1-5

WORD	SUBJECT 1	SUBJECT 2	SUBJECT 3	SUBJECT 4	SUBJ. 5
CHORD	/	/	/	/	/
ACHE	/	/	/	/	/
DEPOT	/	X	/	/	/
AISLE	/	/	/	/	/
BOUQUET	/	/	X	/	/
PSALM	/	/	X	/	/
CAPON	/	/	/	/	/
DENY	/	/	X	/	/
NAUSEA	/	/	/	/	/
DEBT	/	/	/	/	/
COURTEOUS	/	/	/	/	/
RAREFY	/	/	/	/	/
EQUIVOCAL	/	/	X	X	/
NAIVE	/	/	/	/	/
CATACOMB	/	/	/	/	/
GAOLED	/	/	X	/	/
THYME	/	/	X	/	/
HEIR	/	/	/	/	/
RADIX	/	X	X	/	/
ASSIGNATE	/	/	/	/	/
HIATUS	/	/	/	/	/
SUBTLE	/	/	/	/	/
PROCREATE	/	/	/	/	/
GIST	/	/	/	/	/
GOUGE	/	/	X	/	/
SUPERFLUOUS	/	X	/	/	/
SIMILE	/	/	/	/	/
BANAL	/	/	/	/	/
QUADRUPED	/	/	/	/	X
CELLIST	/	/	X	/	/
FACADE	/	/	X	/	/
ZEALOT	/	X	X	/	/
DRACHM	X	X	X	X	X
AEON	/	/	/	/	X
PLACEBO	/	/	X	/	/
ABSTEMIOUS	/	X	X	/	X
DETENTE	/	X	X	/	/
IDYLL	X	X	/	/	/
PUERPERAL	X	X	X	X	X
AVER	X	X	/	/	X
GAUCHE	/	X	/	X	X
TOPIARY	/	X	X	X	/
LEVIATHAN	/	X	X	X	X
BEATIFY	/	X	X	X	X
PRELATE	X	X	X	X	X
SIDEREAL	X	X	X	X	X
DEMESNE	X	X	X	/	X
SYNCOPE	X	X	X	X	X
LABILE	/	X	X	X	X
CAMPANILE	/	X	X	X	X

NB / = Correct Word Pronunciation X = Incorrect Word Pronunciation

Appendix 2b: Scoring Template for subjects 6-10

WORD	SUBJECT6	SUBJECT 7	SUBJECT 8	SUBJECT 9	SUBJ. 10
CHORD	/	/	/	/	/
ACHE	/	/	/	/	/
DEPOT	/	/	/	/	/
AISLE	/	/	/	/	/
BOUQUET	/	/	/	/	/
PSALM	/	/	/	/	/
CAPON	/	/	/	/	/
DENY	/	/	/	/	/
NAUSEA	/	/	/	/	/
DEBT	/	/	/	/	/
COURTEOUS	/	/	/	/	/
RAREFY	/	/	/	/	/
EQUIVOCAL	/	/	/	/	/
NAIVE	/	/	/	/	/
CATACOMB	/	/	/	/	/
GAOLED	/	X	/	X	/
THYME	/	X	/	/	/
HEIR	/	/	/	/	/
RADIX	X	X	/	X	/
ASSIGNATE	/	/	/	/	/
HIATUS	/	/	/	/	/
SUBTLE	/	/	/	/	/
PROCREATE	/	/	/	/	/
GIST	/	/	/	/	/
GOUGE	/	/	/	/	/
SUPERFLUOUS	/	/	/	X	/
SIMILE	/	/	/	/	/
BANAL	/	/	/	/	/
QUADRUPED	/	X	X	X	X
CELLIST	/	/	/	/	/
FACADE	/	/	/	X	/
ZEALOT	/	/	X	X	/
DRACHM	/	X	X	X	X
AEON	/	/	/	/	X
PLACEBO	/	/	/	X	/
ABSTEMIOUS	/	X	X	X	X
DETENTE	/	/	X	X	/
IDYLL	/	/	X	X	/
PUERPERAL	/	X	/	X	X
AVER	X	X	X	X	X
GAUCHE	/	/	X	X	X
TOPIARY	/	X	X	X	/
LEVIATHAN	X	/	/	X	X
BEATIFY	/	X	X	X	X
PRELATE	X	X	X	X	X
SIDEREAL	X	X	X	X	X
DEMESNE	/	X	X	X	X
SYNCOPE	X	X	X	X	X
LABILE	X	/	X	X	X
CAMPANILE	X	X	X	X	X

NB / = Correct Word Pronunciation X = Incorrect Word Pronunciation

Appendix 3: NART word variation score sheet for psychology participants for stage 1

WORD	WORD 1	WORD 2	WORD 3	WORD 4	WORD 5
CHORD					
ACHE					
DEPOT					
AISLE					
BOUQUET	1	2	3		
PSALM					
CAPON					
DENY					
NAUSEA	1	2			
DEBT					
COURTEOUS					
RAREFY					
EQUIVOCAL					
NAIVE					
CATACOMB					
GAOLED					
THYME					
HEIR					
RADIX					
ASSIGNATE					
HIATUS					
SUBTLE					
PROCREATE					
GIST					
GOUGE					
SUPERFLUOUS	1	2			
SIMILE					
BANAL					
QUADRUPED					
CELLIST					
FACADE					
ZEALOT					
DRACHM					
AEON					
PLACEBO					
ABSTEMIOUS					
DETENTE					
IDYLL	1	2			
PUERPERAL					
AVER					
GAUCHE					
TOPIARY					
LEVIATHAN					
BEATIFY					
PRELATE					
SIDEREAL					
DEMESNE	1	2			
SYNCOPE					
LABILE					
CAMPANILE	1	2			

Appendix 4a: Score sheet used by psychologists to rate the 10 participants (1-5) for NART errors

WORD	SUBJECT 1	SUBJECT 2	SUBJECT 3	SUBJECT 4	SUBJ. 5
CHORD					
ACHE					
DEPOT					
AISLE					
BOUQUET					
PSALM					
CAPON					
DENY					
NAUSEA					
DEBT					
COURTEOUS					
RAREFY					
EQUIVOCAL					
NAIVE					
CATACOMB					
GAOLED					
THYME					
HEIR					
RADIX					
ASSIGNATE					
HIATUS					
SUBTLE					
PROCREATE					
GIST					
GOUGE					
SUPERFLUOUS					
SIMILE					
BANAL					
QUADRUPED					
CELLIST					
FACADE					
ZEALOT					
DRACHM					
AEON					
PLACEBO					
ABSTEMIOUS					
DETENTE					
IDYLL					
PUERPERAL					
AVER					
GAUCHE					
TOPIARY					
LEVIATHAN					
BEATIFY					
PRELATE					
SIDEREAL					
DEMESNE					
SYNCOPE					
LABILE					
CAMPANILE					

Appendix 4b: Score sheet used by psychologists to rate the 10 participants (6-10) for NART errors

WORD	SUBJECT 6	SUBJECT 7	SUBJECT 8	SUBJECT 9	SUBJ. 10
CHORD					
ACHE					
DEPOT					
AISLE					
BOUQUET					
PSALM					
CAPON					
DENY					
NAUSEA					
DEBT					
COURTEOUS					
RAREFY					
EQUIVOCAL					
NAIVE					
CATACOMB					
GAOLED					
THYME					
HEIR					
RADIX					
ASSIGNATE					
HIATUS					
SUBTLE					
PROCREATE					
GIST					
GOUGE					
SUPERFLUOUS					
SIMILE					
BANAL					
QUADRUPED					
CELLIST					
FACADE					
ZEALOT					
DRACHM					
AEON					
PLACEBO					
ABSTEMIOUS					
DETENTE					
IDYLL					
PUERPERAL					
AVER					
GAUCHE					
TOPIARY					
LEVIATHAN					
BEATIFY					
PRELATE					
SIDEREAL					
DEMESNE					
SYNCOPE					
LABILE					
CAMPANILE					