Research Article

The National Adult Reading Test (NART) – A preliminary validation study on a Maltese tertiary education population

Anna Girard, Gertrude Rapinett and Richard Muscat
Laboratory of Behavioural Neuroscience, Department of Biomedical Sciences, University of Malta, Msida MSD06, MALTA.

Summary: *The popularity of the National Adult Reading Test (NART) and the North American Adult Reading Test (NAART) as measures of pre-morbid intellectual ability and as accurate estimates of IQ in healthy samples evoked an interest to conduct this preliminary validation study to assess how a Maltese tertiary education population faired on these tests. The National Adult Reading Test-revised (NART-R; Crawford, 1990) or the North American Adult Reading Test (NAART; Blair & Spreen, 1989) was administered on a randomly allocated sample of Maltese tertiary education students and graduates (N=50). The Extended River Mead Behavioural Memory Test (ERBMT) was used as a baseline test to allow for comparisons between the two groups. The results yielded a significant difference in performance between the two NART/NAART groups with a greater number of errors resulting from the NAART group. Comparisons on the overall performance between groups on the ERBMT and on one of the subtests of the ERBMT reflecting semantic memory, yielded no significant difference, indicating that the NART/NAART performances were not a result of pre-existing intellectual group differences. Conversion of NART/NAART scores into WAIS-r IQ's revealed a lower mean IQ than that estimated for tertiary education students or graduates. This observation raises the question of whether the two forms of NART correctly reflect the performance capacity of Maltese graduates and students.*

Keywords: National Adult Reading Test

Introduction

The validity of a test concerns what the test measures and how well it does so. The validity of a test must be established with reference to the particular use for which it is being considered. In the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1985), the specific procedures for determining test validity are grouped under three categories, namely content-related, criterion-related and construct-related. Messick, (1980b) has argued convincingly that the term validity, insofar as it designates the interpretive meaningfulness of a test, should be reserved for construct validity.

When carrying out a validation study on a test that has been standardized on a different population one cannot dismiss the cross-cultural element involved in validating the test for a new population. Examination of the construct validity of a test provides evidence concerning the appropriateness and fairness of the *use* of the test. Comparability of factor analysis results for different groups and the degree to which the results of the factor analysis are consistent with the major scores and common interpretations of the test are necessary conditions for fairness in use of the test with culturally diverse persons. Indeed, if a test is not measuring the same underlying abilities or if the commonly used scores from the test represent varying abilities depending on group membership, then use of the test with culturally different persons is probably inappropriate and unfair, and the predictive validity of the test is likely to be lower for specific groups (Hambleton, 1994).

The attainment of equivalent measures is perhaps the most central issue in cross-cultural/language comparative research. If the basis of comparison is not equivalent across different groups, then valid comparisons across these groups cannot be made. Consonant with the unified conceptualization of validity, assessment bias is regarded as differential construct validity that is addressed by the question: To what extent is the assessment task measuring the same construct and hence has similar meaning for different populations? The presence of bias invalidates score inferences about target constructs because of irrelevant, non-target constructs that affect performance differently across groups (Messick, 1989). These irrelevant constructs are related to characteristics such as gender, ethnicity, race, linguistic background, socioeconomic status (SES), or other conditions that define the groups.

The National Adult Reading Test (NART)

The National Adult Reading Test (NART) is a highly respected measure of pre-morbid intelligence and gives an accurate measure of IQ by assessing the ability to read non-phonetic words – an intellectual function which remains intact after dementia, strokes or head injury. The development of the NART was based on the finding that reading ability is highly correlated with general IQ in the normal population but is maintained at or near its pre-morbid level in patients with dementia.

Since it's publication in 1982, the information about the NART's reliability and validity as a measure of current and premorbid intelligence has become more available.

It is a test that can be administered by both experienced as well as inexperienced clinicians (O'Carroll, 1987; Schlosser & Ivison, 1989), and has shown to be a valid measure of general intelligence in the normal population (Crawford et. al., 1989b) as well as having a potentially wide range of applicability in organic and functional disorders (Nelson & O'Connell, 1978).

Some studies have indicated that NART scores may be only relatively resistant to the effects of progressive dementia in organic disorders (Stebbins et. al., 1990b; Grober & Sliwinski, 1991). Other studies have reported no apparent effect on reading ability in a range of dementing conditions, including dementia of Alzheimer type (DAT), multi-infarct dementia, alcoholic dementia, head injury and AIDS (Crawford et. al., 1988a) As a measure for detecting and assessing the extent of intellectual deterioration, the studies reviewed above suggest that the NART may underestimate premorbid IQ levels only in the more severely demented subject or in those subjects with a more pronounced language deficit and there is no evidence to suggest that the NART does not give a true estimate of premorbid IQ in mildly dementing subjects (Nelson & Willison, 1991). The NART has achieved popularity as a measure of pre-morbid intellectual ability based on the premise that pronunciation of irregular words is unaffected in many clinical disorders and that performance is highly correlated with general intellectual ability (O'Carroll et. al., 1992). Rather, the most common problem with the NART seems to revolve around its insensitivity in estimating intelligence levels above the normal/bright range (Nelson & Willison, 1991).

The original NART standardization study used the Wechsler Adult Intelligence Scale (WAIS) as the test of intelligence; hence, the IQ equivalents of NART scores are WAIS IQ's. With the ever-increasing popularity of the Wechsler Adult Intelligence Scale – Revised (WAIS-R), it became desirable to restandardise the NART against the WAIS-R so that NART scores could be converted directly to WAIS-R IQ's and give a more accurate indication of intelligence level relative to today's population. The restandardisation was based on 182 people aged 18 to 70 years, including volunteers, non-neurological hospital in-patients and other normal subjects. The ability of the NART (Nelson and Willison, 1991) and the revised NART (NART-r), (Crawford et. al., 1990) to estimate IQ was examined in 47 healthy subjects using the Wechsler Adult Intelligence Scale-Revised (WAIS-R) scores as the criterion. The NART-r showed to have significantly higher correlations with Full Scale and Verbal IQ than the NART. Published reports indicate that the NART-r can be used reliably with people aged up to 84 years (Nelson & Willison, 1991).

To examine the relationship between NART performance and demographic variables, Nelson's (1982) original report of the NART's split-half reliability was re confirmed in a study where performance and demographic variables were examined. This study, which used subjects free of neurological or psychiatric disorder, reported that the NART estimated IQ was significantly correlated with education, social class and age (Crawford et al., 1988b)

The NART was originally designed to provide a means of estimating the pre-morbid intelligence levels of adults suspected of intellectual deterioration. Performance depends more on previous knowledge than on current cognitive capacity (Nelson & O'Connell, 1978). The value of the test lies in the high correlation between reading ability and intelligence in the normal population (Crawford et. al., 1988b). Nelson developed the test in England for use with the WAIS. Recently, in 1991, Nelson and Willison restandardised the test on a British sample making it possible to convert NART-R scores directly to WAIS-r scores (Nelson and Willison, 1991).

Description of the NART
(National Adult Reading Test)

A list of 50 words printed in order of increasing difficulty is read aloud by the examinee. Each word is relatively short and irregular in terms of common rules of pronunciation, in order to minimize the possibility of reading by phonemic decoding rather than word recognition. From the reading error scores obtained, verbal, performance and full-scale IQ scores can be predicted to approximate closely the pre-morbid IQ level. Restandardised (based on 182 people aged 18-70yrs), and drawing on many studies of reliability and validity published over almost a decade, the test provides:

- Predicted scores for the WAIS-R
- Improved large print materials for use with people who are partially sighted, mentally frail of who have dementia.
- Improved sensitivity at higher IQ levels.
- A range of new validation studies

Blair and Spreen (1989) modified the test for use with American populations. The North American Adult Reading Test (NAART) has been validated against the WAIS-R. The NAART consists of 61 irregular words printed in 2 columns, and is administered and corrected in the same way as the English NART-R. In a study comparing the NAART and the Wide Range Achievement Test – Revised (WRAT-R) on neurologically impaired patients, Johnstone et. al. (1996) reported that although both tests were equivalent and accurate estimates of average verbal IQ level, the WRAT-R had superior normative data and a less restricted range and standard deviations equal to that of the WAIS-R than the NAART (Johnstone et. al. 1996).

The Extended Rivermead Behavioural Memory Test (ERBMT)

The Extended Rivermead Behavioral Memory Test (ERBMT) is a test, which is used to assess every day memory. Most clinical memory tests are extensions of laboratory tests, and do not as such map directly onto memory problems encountered by patients in their everyday lives. The Rivermead Behavioral Memory Test (RMBT) was devised to solve this problem (Wilson, B.A. et. al., 1989). It consists of a number of subtests each attempting to provide an objective measure of one of a range of everyday memory problems. It has been validated using the observation of memory lapses in a sample of 80 brain-damaged patients observed over an average of 55 hours each. It proved to have a high correlation with observed lapses and to have high inter-rater reliability. Four parallel version of the test are available and are applicable to a wide range of environmental settings, making the RBMT a short, reliable and valid test of everyday memory problems. The RBMT, however, was originally designed as a screening test and thus is insufficiently sensitive to detect mild deficits, whether due to brain damage or to the introduction of a drug or stressor. Therefore, the RBMT was extended to provide a sensitive measure of memory within the normal range.

The Extended Rivermead Behavioral Memory Test (ERBMT) increases the level of difficulty by doubling the amount of material to be remembered, by combining material from Forms A and B and forms C and D of the original test to produce 2 parallel versions of the new extended test. The sensitivity of the ERBMT was assessed by comparing the performance of a middle-aged man and an elderly group of normal subjects, who would be expected to show modest differences in memory performance. The subtests varied in their sensitivity to this small age difference, but when performance was assessed in terms of scale scores that allow an overall combined measure of memory performance to be calculated, the test proved sensitive and free of ceiling and floor effects. (Wilson, B.A. et al., 1989). It has been suggested that the ERBMT provides a promising measure of everyday memory in normal adults. With regards to ecological validity of the test, it has been argued that the ERBMT has higher ecological validity than other memory tests (De Wall, C. et. al., 1994), as it was initially based on contrasting groups for whom everyday memory problems were or were not prominent and subsequently was validated against many hours of careful observation to finally being used effectively to predict the capacity of patients to cope independently. What may be arguable is the ecological validity of each of the subtests to detect the milder deficits in everyday memory performance, a factor that could extend the use of the test across an even wider range of situations.

Materials and methods

Participants
Participants (N=50) aged between 18 and 37 were selected for the study. All participants were recruited randomly by word of mouth over a period of 2 months. They were selected so that personal characteristics (e.g. age, gender and education) in each of the experimental groups were matched. The participants used for this study were either university students or university graduates. Criteria of exclusion included non-English speakers.

Materials
The Extended Rivermead Behavioural Memory Test (ERBMT) was used as a baseline memory test to allow for comparison between the 2 groups. This test fulfills criteria required for the experiment and is a reasonably sensitive, validated and culture-fair everyday memory test (de Wall, C., Wilson, B., & Baddeley, A. 1994).

Either the National Adult Reading Test NART (UK) or the North American Reading Test NAART (US) was administered to each participant either directly before or directly after the administration of the ERBMT.

Test Procedure
Two experimental groups were randomly assigned to NART (UK), (n=25) and NAART (US), (n=25) conditions. Each experimental group was further divided in two sub groups according to the sequence of administration of tests.

Exp. group 1. (NART) (UK)		Exp. Group 2. (NART) (US)	
Group A	*Group B*	*Group A*	*Group B*
NART	ERBMT	NART	ERBMT
ERBMT	NART	ERBMT	NART

The four groups were balanced to eliminate performance variance as a result of motivation and/or subjective difficulty on either one of the tests. Age and gender were also balanced across the different conditions. Before the commencement of tests each participant was debriefed on the procedure of the tests and asked to read and sign a Consent For. The duration of the tests was approximately 40 minutes per person.

Scoring
The results of the tests were scored using the standardized scoring sheets accompanying each test. The raw data was processed and analysed using SPSS version 10.0. NART/NAART scores were converted into WAIS-R full and verbal IQ scores (Wechsler, D. 1981).

Results
Analyses of the results demonstrated that the sample population (N=50) used for this study performed better on the NART (National Adult Reading Test) then on the NAART (North American Adult Reading Test). The

descriptive statistics for the participants are given in Table 1.1 and Figure 1 below.

Table 1.1: Descriptive statistics of subjects used in the validation study (N=50)

	N	Min.	Max.	Mean	Std. Dev.
NART	25	11	32	19.88	5.11
NAART	25	15	38	23.64	5.85

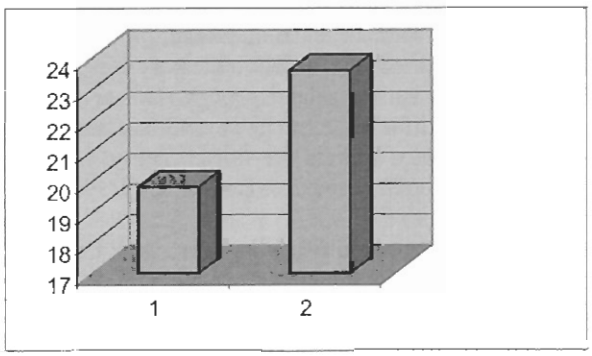

Fig.1 Mean error scores of Group 1 (NART) 19.88 compared to Group 2 (NAART) 23.64

The error scores on both the NART and NAART groups was obtained and converted into the WAIS-R predicted Full Scale IQ (Table 1.2)

Table 1.2: The WAIS-R Full Scale and Verbal IQ's predicted from range of errors made.

	Error Scores	Means
Range IQ	**117 – 91**	**106.1**
NART	11 – 32	20
Range IQ	**112- 84**	**101.0**
NAART	15 – 38	24

Comparison of NART and NAART means using an Independent Samples T-Test showed that there was a significant difference in errors made between the two experimental groups t (48) = -2.42, $p<.01$. In order to verify that this difference was not attributed to IQ differences between the two groups, an Independent Samples T-Test was administered between the two groups on their performance on the ERBMT baseline test. The results of this test showed that there was no significant difference between the 2 groups on the ERBMT $t(48) = 1.06$, $p>.05$, thus eliminating intelligence level as possible explanation for the significant difference between the two groups.

Further analysis on one of the sub-tests of the ERBMT; the 'immediate' and 'delayed' recall of a 'story' which taps the use of semantic memory was carried out between the NART and NAART groups. The Mean scores on performance in this subtest were compared between the groups (Table 1.3) and the results from an Independent Samples T-Test used for this analysis showed no significant difference between the two groups, both on 'immediate' recall of the story $t(48) = -0.169$ $p>0.05$, as well as on 'delayed' recall $t(48) = -0.926$ $p>0.05$.

Table 1.3: Comparison of Mean scores from the 2 groups on ERBMT Story Recall

		N	Mean	Std. Dev.	Std. Error Mean
Story 1	NART	25	2.24	0.88	0.18
	NAART	25	2.28	0.79	0.16
Story 2	NART	25	1.6	0.87	0.17
	NAART	25	1.8	0.65	0.13

Story 1 = Immediate Recall; Story 2 = Delayed Recall

Furthermore, scores obtained from the NART/NAART were correlated with scores obtained from both the 'immediate' as well as the 'delayed' recall of the ERBMT story. A Pearson's Correlation between NART/NAART scores and 'immediate' recall scores showed no significant difference $r = -0.18$, $p>.05$. Similarly, the NART/NAART and 'delayed' story correlation also yielded no significant difference $r = -0.06$, $p>.05$.

A further analysis to check the validity of transforming the NART/NAART scores into WAIS-R IQ scores was made. A correlation was made between NART/NAART converted Full and Verbal IQ scores and ERBMT overall performance. A Pearson's Correlation was administered and yielded a significant correlation between both Full IQ scores and ERBMT overall scores, $r = .436$, $p<.01$ as well as Verbal IQ scores and ERBMT overall scores, $r = .436$, $p<.01$.

An Independent Samples T-Test was carried out to see whether there were any gender differences in performance on the NART/NAART. Analysis of the results yielded no significant difference $t(48) = -0.382$, $p>.05$.

Discussion

The NART group in the study committed less pronunciation errors than the group allocated to the NAART, indicating that performance ability on the NART was greater than that for the NAART. This observation was further confirmed by the results obtained from comparisons on performance between the two groups on the ERBMT baseline memory test and on

the ERBMT semantic memory sub-test, which as explained previously, yielded no significant difference between the two groups on either of the comparisons.

The sample used in both groups for this study was homogeneous with regards to age, sex, education and social background; therefore none of the above confounding variables could be attributed to the observed differences in performance between the two groups. With regards to age and sex, Crawford et. al., (1988b), have reported that the two variables have little effect on performance and that an age-related increase in *correct* NAART scores only appears to emerge when a wide range of persons are studied (well-educated healthy individuals aged 16-84). In this study, the range of age of persons used was between 18 years and 37 years, also the educational background of the participants was similar. However, the type of course that a person had frequented or was reading at the University might have had an influence to prior knowledge of certain words, for example the word *leviathan* for philosophy students.

A relevant factor which merits mention with regards to the differences in performance observed between the two NART groups is that Maltese schooling is based on a British style of education and therefore the Maltese students' exposure to English words is more congruent with the NART than the NAART. This observation could contribute to one of the explanations related to the better performance achieved from the NART as compared to that achieved from the NAART.

Comparisons of the overall performance on the ERBMT between the NART and NAART groups were made to justify the results observed in the NART/NAART tests. The ERBMT has been validated across the normal range and has shown to be a reliable and sensitive measure of memory within this range (de Wall et. al., 1994). The results from these comparisons showed no significant differences in the performance levels on the ERBMT between the two NART and NAART groups and therefore the differences observed in performance between the NART and the NAART cannot be attributed to intellectual differences between the two groups. The fact that both groups also showed no significant differences in the semantic memory 'immediate' and 'delayed' *story recall* of the ERBMT further highlighted this assumption.

A point of interest with regards to this study is that the average predicted Full Scale IQ obtained from performances in the NART/NAART was that of 100. Although this is an average score it is lower than the estimated IQ for tertiary education students or graduates, which ranges between 110 and 120. This observation raises the question of whether the two forms of NART correctly reflect the performance capacity of Maltese graduates and students. A sample of 50 persons is too small to arrive at any solid conclusions with regards to this observation; however, this could be a starting point for a series of future validation studies of the NART/NAART on the Maltese population using a larger sample of same education / social background participants which, depending on the type of results obtained, may or may not lead to further studies on a wider more heterogeneous sample. The interesting point in question that emerges from this study is whether or not the NART/NAART can be validated and used reliably on the Maltese population. Furthermore, the results obtained from this study could be indicative as to which test, the NART or the NAART could be more appropriate for the Maltese population, if at all.

References

Crawford, JR, Parker, DM, Besson, JAO (1988a). Estimation of premorbid intelligence in organic conditions. *British Journal of Psychiatry*, **153**, 178-181.

Crawford, JR, Stewart, LE, (1988b). Garthwaite, PH, Parker, DM, Besson, JA. The relationship between demographic variables and NART performance in normal subjects. *Br J Clin Psychol* **27** (Pt 2): 181-182.

Crawford, JR, Stewart, LE, Cochrane, RHB, Parker, DM, Besson, JAO (1989a). Construct validity of the National Adult Reading Test: A factor analytic study. *Personality and individual differences*, **10**, 585-587.

Crawford, JR, Allan, KM, Cochrane, RHB, Parker, DM (1990). Assessing the validity of the NART-estimated premorbid IQs in the individual case. *British Journal of Clinical Psychology* **29**, 435-436.

De Wall, C, Wilson, BA, Baddeley, AD (1994). The Extended Rivermead Behavioural Memory Test: A measure of everyday performance. *In Memory*, MRC Applied Psychology Unit, Cambridge, 149-166.

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

Hambleton, RK (1994). Guidelines for adapting educational and psychological test: A progress report. *European Journal of Psychological Assessment*, **10**, 229-244.

Johnstone, B, Callahan, CD, Kapila, CJ, Bouman, DE (1996). The comparability of the WRAT-R reading test and the NAART as estimates of premorbid intelligence in neurologically impaired patients. *Archives of Clinical Neuropsychology*, **11**, 513-519.

Messick S (1989). Validity. In Linn R (Ed.), *Educational Measurement*, New York: Macmillan Publishing Company, 221-262.

Nelson, HE, Willison, JR, (1991). National Adult Reading Test (NART): *Test Manual. Second Edition.* Windsor, UK: NFER Nelson.

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

O'Carroll, R (1987). The Inter-rater reliability if the National Adult Reading Test (NART): a pilot study. *Br J Clin Psychol*: 229-230.

O'Carroll, R, Walker, M, Dunan, J, Murray, C, Blackwood, D, Ebmeir, KP, Goodwin, GM (1992). Selecting controls for schizophrenia research studies: the use of the NART is a measure of premorbid ability. *Schizophrenia Research*, **8**, 137-141.

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.

Stebbins, GT, Wilson, RS, Gilley, DW, Bernard, BA, Fox, JH (1990b). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *The Clinical Neuropsychologist*, **4**, 18-24.

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

Wilson, B, Cockburn, J, Baddeley, A (1985). *The Rivermead Behavioural Memory Test*. Bury St. Edmunds, England: Thames Valley Test Company.