British journal of Clinical Psychology (1989), **28,** 267-273 *Printed in Great Britain*

Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension

J. R. Crawford* and D. M. Parker

Department of Psychology, University of Aberdeen, Aberdeen AB9 2UB, UK

L. E. Stewart, J. A. O. Besson and G. De Lacey

Department of Mental Health, University of Aberdeen

Nelson (1982) presented regression equations for the prediction of Wechsler Adult Intelligence Scale (WAIS) IQ from performance on the National Adult Reading Test (NART). In a cross-validation sample *(n* = 151) these equations predicted 66, 72 and 33 per cent of the variance in WAIS Full Scale, Verbal and Performance IQ respectively. There were no ceiling or floor effects in the relationship between NART performance and WAIS IQ despite the wide IQ range of the sample. The standardization and cross-validation samples were combined *(n* = 271) to generate new regression equations. These equations should be used in preference to the original equations as they are based on a larger sample with a wider IQ and age range. Combining NART and Schonell Graded Word Reading Test errors did *not* improve IQ prediction in poor readers. A detailed examination of the NART's test-retest and inter-rater reliability was also conducted.

The National Adult Reading Test (Nelson, 1982) consists of 50 short words of irregular pronunciation (e.g. *deny. prelate).* Nelson (1982) presented regression equations for the prediction of Wechsler Adult Intelligence Scale (Wechsler, 1955) performance from NART errors. These equations can be used to provide an estimate of premorbid IQ as NART performance is largely resistant to the effects of neurological and psychiatric disorder (e.g. Crawford, Besson, Parker, Sutherland & Keen, 1987; Crawford, Parker & Besson, 1988; O'Carroll, Baikie & Whittick, 1987).

The first aim of the study described here was to examine the predictive accuracy of Nelson's (1982) regression equations in a cross-validation (CV) sample of healthy, normal subjects. Subsidiary aims were to determine *(a)* if there were any ceiling or floor effects in the relationship between NART performance and IQ, and *(b)* whether combining NART and Schonell Graded Word Reading Test errors increased the predictive accuracy of IQ estimates in poor readers.

* Requests for reprints. An extended version of the paper is also available from the authors.

When regression equations are applied to CV samples there is normally shrinkage in the amount of variance they account for in the criterion variable. If this shrinkage is not excessive, regression equations based on the combined standardization and CV samples should be generated as they will have more stability than the original equations (e.g. Pedhazur, 1982). This procedure was to be carried out for the NART, provided that the foregoing precondition was met.

A second aim of the study was *(a)* to determine if the NART has an acceptable test-retest reliability and *(b)* to determine whether or not a statistically and clinically significant practice effect occurs on retesting. The third aim was to investigate issues associated with the NART's inter-rater reliability (IRR). It was intended *(a)* to attempt to replicate O'Carroll's (1987) report of a high IRR in a larger sample (with a wider range of NART performance), *(b)* to determine whether the inclusion of raters with no previous experience of the NART lowers IRR, *(c)* to investigate whether particular words produce a disproportionately high degree of inter-rater disagreement and *(d)* to determine if there were -any marked differences in the strictness of scoring between raters.

Method

Subjects in the three studies described below were recruited from a wide variety of sources, i.e. local and national companies, local clubs (e.g. pensioner clubs, angling clubs), community centres, etc. Most received a small honorarium. All were free of neurological, psychiatric or sensory disability.

Cross-validation study

The sample consisted of 151 subjects (79 males, 72 females). Mean age was 42.0 years (SD = 17.1) with an age range of 16-88 years and mean years of education were 12.6 (SD = 3.1). The WAIS, NART and Schonell Graded Word Reading Test (Schonell, 1942) were administered and scored according to standardized procedures. NART error scores were entered into Nelson's (1982) regression equations to obtain NART estimated Full Scale (FSIQ), Verbal (VIQ) and Performance IQ (PIQ).

Reliability studies

For the test-retest study (TRS) 61 subjects were administered the NART and retested after 10 days (±1 day). Mean age was 37.1 years (SD = 12.2). For the inter-rater study (IRS) the NART performance of 40 subjects was audiotaped. Ten clinical psychologists acted as raters: five routinely administered the NART (experienced raters) whereas the remaining five had no previous experience with the test (naive raters). Raters were asked to rate NART errors using the pronunciation guide provided in the NART manual (Nelson, 1982). They were asked not to confer with colleagues over pronunciation.

Results

Cross-validation study

Mean FSIQ (111.8 ±12.7) and mean NART errors (20.5 + 5.8) were similar to the standardization sample. However, the present sample had a wider IQ range (75-140 vs. 86-128) and a wider age range (16-88 vs. 20-70). The three WAIS scales were regressed on their corresponding NART-estimated IQ. The results of this

procedure are presented in Table 1. It can be seen that in the present sample Nelson's (1982) regression equations predicted 66, 72 and 33 per cent of the variance in WAIS FSIQ, VIQ and PIQ, respectively. To investigate whether there were any non-linear trends in the relationship between NART performance and IQ, quadratic and cubic functions of NART error score were added to the regression models described above. These failed significantly to increase predicted IQ variance. This result is encouraging as it indicates that, despite the wide IQ range examined, there were no appreciable ceiling or floor effects in the relationship between NART performance and IQ.

Table 1. Percentage of WAIS variance accounted for by the NART in the standardization sample and present cross-validation sample (CVS)

	Full Scale IQ	Verbal IQ	Performance IQ
Standardization sample ($n = 120$)	55 (7.6)	60 (7.6)	32 (9.4)
CVS ($n = 151$)	66 (7.4)	72 (7.5)	33 (9.5)
CVS – prorated WAIS	60 (7.9)	64 (8.1)	30 (10.4)

Note. Standard error of estimate in parentheses.

It can be seen from Table 1 that the NART equations accounted for *more* IQ variance in the CV sample than in the standardization sample. This is partly attributable to the fact that WAIS IQs in the present sample were derived from a full, rather than prorated, WAIS. When WAIS IQs were prorated with the subtests used in the standardization sample there was a decline in the variance accounted for although, with the exception of PIQ, this was still greater than in the standardization sample (see Table 1).

As noted above, it is highly desirable to generate regression equations from combined standardization and CV samples provided that an excessive shrinkage in the variance does not occur in the CV sample. As this precondition was clearly met, the standardization and CV samples were combined *(n* = 271). Regression equations were built for the prediction of (prorated) FSIQ, VIQ, and PIQ from NART error score. These equations predicted 57, 63 and 31 per cent of the variance in FSIQ, VIQ and PIQ, respectively, and are presented below with their corresponding standard errors of estimate:

predicted FSIQ = 128.50-0.84 (NART errors) standard error of estimate = 7.68
predicted VIQ = 129.91-0.94 (NART errors) standard error of estimate = 7.73
predicted PIQ = 122.85-0.60 (NART errors) standard error of estimate = 9.57

The distribution of positive predicted minus obtained IQ discrepancies in the combined sample are presented in Appendix 1. In clinical practice this table can be used to assess the probability of a particular size of discrepancy occurring in the normal population.

Nelson (1982) presented regression equations based on combined NART and Schonell errors which she advised should be used to predict WAIS IQ in poor readers (operationally defined as subjects with a NART error score greater than 40).

To examine whether this combination would provide a more accurate estimate in such subjects, regression equations, based on NART and Schonell errors, were built from the combined sample. Seven subjects (two in the standardization sample, five in the CV sample) had NART error scores greater than 40. Comparison of the NART- and NART + Schonell-estimated IQs revealed that the NART provided a more accurate estimate. The regression equations based on NART + Schonell errors systematically underestimated WAIS IQ in this subgroup to the extent that, in the sample as a whole, there was a significant non-linear component in the relationship between predicted and obtained IQ (i.e. adding polynomial functions of NART + Schonell errors to the equation significantly increased predicted IQ variance).

Test-retest study

In the TRS group, NART errors at first testing ranged between 3 and 42 with a mean of 17.98 (10.8). The correlation (Pearson product moment) between NART errors at test and retest was 0.98 ($P < 0.01$) indicating that the NART has an extremely high test-retest reliability. Although there was a significant decrease in NART errors at retesting (/ = 2.30, $P < 0.05$), the mean change was less than one NART error (17.98 vs. 17.26) suggesting that practice effects are of little practical significance.

Inter-rater study

The Pearson product moment correlations between all possible pairs of *experienced* raters ranged between 0.96 and 0.98. The same range of coefficients was obtained when the correlations between *all* raters were examined. The study described here therefore replicates O'Carroll's (1987) report of high IRR in a larger sample with a wider range of NART performance and also indicates that previous experience with the NART is not necessary in order to score it reliably. Mean rated NART errors for the individual raters ranged from 15.3 to 19.3 with an ove-rall mean of 17.3. To determine if raters differed significantly in the strictness/leniency of their scoring a two-way analysis of variance was performed with subjects and raters as factors. There was, as expected, a highly significant difference between subjects in NART performance ($F - 342.04$, d.f. = 39,351, $P < 0.0001$). There was also a highly significant difference between raters ($F = 26.11$, d.f. = 9,351, $P < 0.0001$). Thus, raters differed significantly in the strictness with which they scored the NART.

Examination of the percentage agreement between raters for individual NART words revealed that 82 per cent of words had an agreement rate of 90 per cent or greater while 64 per cent had an agreement rate of 95 per cent or greater. The five words with the lowest agreement rates were: *aeon* (71.25 per cent), *puerperal* (80.75 per cent), *aver* (83.25 per cent), *sidereal* (85.25 per cent) and *prelate* (86.0 per cent). Particular care should therefore be taken when recording and scoring these words.

Discussion

Nelson's (1982) regression equations predicted a very substantial proportion of WAIS IQ variance in the CV sample. The results obtained here therefore suggest that

the NART provides a valid estimate of IQ. It should be noted, however, that in the original standardization sample, the CV sample and the combined sample, the NART was a relatively poor predictor of PIQ. Thus, although PIQ tends to be more severely affected than FSIQ and VIQ in many neuropsychiatric conditions, it may be that predicted/obtained PIQ discrepancies have lower discriminative ability than FSIQ or VIQ discrepancies (because relatively large discrepancies in favour of predicted PIQ occur by chance in the normal population). It would be useful to carry out a study employing discriminant function analysis with normal and neuropsychiatric groups to investigate this issue.

The equations derived from the combined samples should be used in preference to the original equations as *(a)* they are based on a larger sample collected from more than one geographical area and *(ti)* the IQ. range and age range of the sample is greater than the original standardization sample.

The findings from the study under report and past research demonstrate that the NART is one of the most reliable tests currently used in psychological practice and research. Split-half reliabilities of 0.90 and 0.93 have been reported (Crawford, Stewart, Garthwaite, Parker & Besson, 1988; Nelson, 1982).

Finally, predictive accuracy is a crucially important determinant of the utility of regression-based methods of estimating premorbid IQ. Other things being equal, the higher the predictive accuracy, the finer the discrimination between impaired and normal intellectual functioning. In the CV sample the predictive accuracy of NART IQ estimates was greater for the full WAIS than the prorated WAIS. In view of this, and since a full WAIS is commonly administered in clinical practice (in which case the prorated equations should not be used), the equations for prediction of full WAIS IQ are presented in Appendix 2 along with their corresponding discrepancy tables. These are necessarily based on the CV sample alone. The results reported here also suggest that, when the NART is restandardized against the WAIS-R (Lea, 1986; Wechsler, 1981), the full version should be administered. They also suggest that *aeon* (which had an agreement rate nearer to chance than perfect agreement) and probably the other individual words listed above, should be replaced during this process.

Acknowledgements

We are grateful to Hazel E. Nelson for allowing us access to the standardization sample data. Thanks also to Steven Bell, Richard Blakey, Helen Burr, Leonara Harding, David Johnson, John Moore, Alison Peaker, Juliana Power and Malcolm McFadyen for acting as raters in the IR study.

References

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., Parker, D. M. & Besson, J. A. O. (1988). The relationship between demographic variables and NART performance in normal subjects. *British Journal of Clinical Psychology,* 27, 181-182. Lea, M. (1986). *A British Supplement to the Wechsler Adult Intelligence Scale-revised.* San Antonio, TX: Psychological Corporation.

Nelson, H. E. (1982). *National Adult Heading Test (NART): Test Manual.* Windsor: NFER-Nelson.
O'Carroll, R. E. (1987). The inter-rater reliability of the National Adult Reading Test (NART): A pilot study. *British journal of Clinical Psychology,* 26, 229-230. O'Carroll, R. E., Baikie, E. M. & Whittick, J. E. (1987). Does the National Adult Reading Test hold in dementia? *British journal of Clinical Psychology,* 26, 315—316. Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research: Explanation and Prediction,* 2nd ed. New York: Holt, Rinehart & Winston.
Schonell, F. (1942). *Backwardness in the Basic Subjects.* London: Oliver & Boyd. Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation. Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale- Revised.* New York: Psychological Corporation.

Received 31 March 1988; revised version received 11 July 1988

Appendix 1

Distribution of positive predicted – obtained IQ discrepancies in normal subjects

Predicted minus obtained discrepancy	% of subjects when FSIQ is predicted	% of subjects when VIQ is predicted	% of subjects when PIQ is predicted
1	48	47	48
2	42	41	44
3	38	37	41
4	34	33	37
5	31	29	32
6	26	25	28
7	20	21	25
8	15	18	23
9	11	14	20
10	9	11	17
11	7	8	15
12	6	6	11
13	6	5	8
14	5	3	7
15	3	2	6
16	3	2	4
17	2	1	4
18	2	1	4
19	1	1	3
20	1	1	2
21	1	1	1
22	1	0	1
23	0	0	0

Note. The figures opposite the discrepancy scores represent the percentages of normal subjects who exhibited that size of positive discrepancy *or larger*. For example, in the case of FSIQ, 2 per cent of subjects exhibited a discrepancy of 10 IQ points and a further 7 per cent of subjects showed a discrepancy greater than 10. Therefore the percentage opposite a discrepancy of 10 is 9 per cent.

Appendix 2

Regression equations for the prediction of the full WAIS from NART errors along with their corresponding predicted minus obtained discrepancy tables

Predicted FSIQ = 131.36 − 0.95 (NART errors) standard error of estimate = 7.37
Predicted VIQ = 134.87 − 1.11 (NART errors) standard error of estimate = 7.48
Predicted PIQ = 122.68 − 0.61 (NART errors) standard error of estimate = 9.54

Predicted minus obtained discrepancy	% of subjects when FSIQ is predicted	% of subjects when VIQ is predicted	% of subjects when PIQ is predicted
1	44	48	45
2	42	42	43
3	36	38	40
4	33	34	35
5	30	32	29
6	24	24	26
7	21	20	23
8	18	16	21
9	12	12	18
10	7	9	17
11	5	7	14
12	3	5	12
13	3	5	10
14	3	2	7
15	3	1	5
16	3	1	5
17	3	1	5
18	3	1	3
19	1	1	3
20	1	1	2
21	1	1	2
22	0	1	2
23	0	0	2
24	0	0	2
25	0	0	2
26	0	0	1
27	0	0	1
28	0	0	1
29	0	0	1
30	0	0	0