

The validity of abbreviated forms of the National Adult Reading Test and Spot-the-Word 2 for estimating full-scale IQ

Ian van der Linde, Lewis Horsman & Peter Bright

To cite this article: Ian van der Linde, Lewis Horsman & Peter Bright (2021): The validity of abbreviated forms of the National Adult Reading Test and Spot-the-Word 2 for estimating full-scale IQ, *Neuropsychological Rehabilitation*, DOI: [10.1080/09602011.2021.1958874](https://doi.org/10.1080/09602011.2021.1958874)

To link to this article: <https://doi.org/10.1080/09602011.2021.1958874>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 29 Jul 2021.



[Submit your article to this journal](#)



Article views: 78



[View related articles](#)



[View Crossmark data](#)

The validity of abbreviated forms of the National Adult Reading Test and Spot-the-Word 2 for estimating full-scale IQ

Ian van der Linde ^{a,c}, Lewis Horsman^{a,b} and Peter Bright ^{a,b}

^aCognition Group, ARU Centre for Mind and Behaviour, Anglia Ruskin University, Cambridge, UK;

^bSchool of Psychology and Sport Science, Anglia Ruskin University, Cambridge, UK; ^cSchool of Computing and Information Science, Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, UK

ABSTRACT

In this study, we validate an earlier proposal for an abridged 17-item National Adult Reading Test (NART) by comparing its performance in estimating full-scale IQ against both the full test and the Spot-the-Word 2 (STW-2) test in a new cohort. We also compare the performance of the 17-item NART to two previous attempts to shorten this test, the Mini-NART and the Short NART. Findings include that NART 17 is numerically stronger and statistically equivalent to the other short variants, the full 50-word NART, and STW-2. Unlike the Short NART, the 17-item NART is usable for participants of all ability levels rather than only those with low reading ability, while offering equally precise premorbid estimates. We also compute that two-thirds of STW-2 is ostensibly redundant for full-scale IQ estimation and we, therefore, propose that, subject to additional verification in an independent sample, an abridged version of this test may also benefit clinical practice.

ARTICLE HISTORY

Received 4 May 2021
Accepted 19 July 2021

KEYWORDS

Premorbid intelligence;
Neuropsychological test
abbreviation; NART; STW-2;
WAIS-IV.

Introduction

Tests requiring oral pronunciation of irregular words remain commonly used by clinicians to estimate the full-scale IQ (FSIQ) of patients prior to brain injury or the onset of a neurological condition. There is good evidence that single word pronunciation knowledge is a well preserved (held) ability across a wide range of conditions (Bright et al., 2002; Bright & van der Linde, 2020; Crawford, 1992; McGurn et al., 2004; O'Carroll, 1995; Sharpe & O'Carroll, 1991), and that the relationship between word reading and intelligence is largely independent of age and social class (Nelson, 1982). Comparison of premorbid IQ estimates

CONTACT Ian van der Linde  ian.vanderlinde@aru.ac.uk  Faculty of Science & Engineering, Anglia Ruskin University, East Road, Cambridge CB1 1PT, UK

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

against objective measures of current IQ on the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008) enables the magnitude of cognitive impairment to be evaluated and is useful for diagnostic and clinical management purposes. A genetic algorithm (GA)-based method that tackles the computationally intractable (NP-hard) problem of identifying optimal subsets of words from premorbid IQ word-reading tests was recently proposed. This method, described in detail in van der Linde and Bright (2018), was demonstrated using the widely used National Adult Reading Test (NART; Nelson, 1982; Nelson & Willison, 1992), for which a subset of 17 of the original 50 words was identified that predict FSIQ as accurately as the entire test. Abbreviated tests have multiple benefits: they reduce the burden on patients and clinicians, help to ameliorate fatigue effects, reduce anxiety and embarrassment for low-scoring participants (Beardsall & Brayne, 1990), and potentially reduce the exposure of patients to clinicians at a time of increased caution due to Covid-19. Cross-validation was used to examine test validity for this earlier finding with $n = 92$ neurologically healthy participants. In the present study, the robustness of the abbreviated 17-word NART (hereafter referred to as NART17) is examined in a confirmatory analysis using a new cohort of participants ($n = 69$), giving 161 participants in total, some 136 more than were used in the development of the NART, and 41 more than were used in its original standardisation (Nelson, 1982). We also compare the predictive accuracy of the NART to Spot-the-Word 2 (STW-2; Baddeley & Crawford, 2012), an alternative 100-item lexical decision test for estimating premorbid IQ and apply our GA procedure to the STW-2 to determine whether an abbreviated version of this test might similarly offer equivalent or enhanced precision in the estimation of WAIS-IV full-scale IQ.

Method

Participants

It is critical to use unimpaired rather than clinical participants to develop and refine hold tests, since it is necessary to measure the relationship between the proposed hold test score and the predicted variable in the normative population (Crawford et al., 1998). 69 participants with no reported history of brain disorder (34 male, 35 female) were recruited by opportunity sampling, ranging in age from 18–67 (mean 38.45, SD 16.52). The recruited participants completed between 10 and 23 years of education (mean 15.67, SD 2.64) and none had previously undertaken any of the tests employed in this study. All participants were UK residents and spoke English as their first language. No neurological condition or learning difficulty likely to affect their performance was declared (self-reported). All participants had normal/corrected-to-normal vision and hearing. Ethical approval was awarded for the study by the relevant

university research ethics panel. Participants consented to participate in writing, and were treated in accordance with the tenets of the Declaration of Helsinki. We report all data exclusions, all manipulations, and all measures in the study.

Apparatus

In addition to all 10 core subtests of the WAIS-IV (Wechsler, 2008), participants completed the NART (Nelson, 1982; Nelson & Willison, 1992) and STW-2 (Baddeley & Crawford, 2012). All tests are the current revisions at the time of writing and were administered in-person. Ordinary procedures for the NART entail the presentation of a sequence of 50 written words to participants, who are asked to pronounce each word. All words in this test are irregular, in that they have atypical grapheme-phoneme relationships and thus prior familiarity is assumed to govern whether they are spoken correctly. For instance, without prior familiarity, the word “pint” might be spoken to rhyme with “mint.” The experimenter evaluates whether pronunciation was correct with the use of a phonetic guide. In the original development of the NART, 140 irregular words were initially identified and tested using $n = 25$ participants (15 “extra-cerebral” inpatients predominantly with spinal cord disorders and peripheral neuropathies, and 10 relatives). Words that were read correctly by all participants were discarded, along with words subjectively judged to be susceptible to guessing and words that would be difficult to evaluate accurately (e.g., that relied upon syllabic stress only). The 50 words chosen were then tested using a new cohort of $n = 120$ extra-cerebral patients (20–70 years of age), with an approximately equal number of each age decade, producing the initial WAIS regression equation.

In this study, the 17 NART words identified in van der Linde and Bright (2018) were presented separately to the remaining (ostensibly redundant) 33 words. The predictive power of the NART17 was compared with the full 50-word NART by pooling the 17 and 33 word results together for each participant. In STW-2, participants are presented with 100 pairs of words and are asked to identify which one of the two is a real word. In this test, participants are simply required to point to the real word (and not the pseudoword) in each word pair, rather than speak it, which has the benefit of being usable in participant groups with speech production difficulties or for whom spoken answers may elicit stress. Hereafter, we refer to NART17, full NART and STW-2 collectively as *word tests*, to distinguish them from the WAIS-IV battery.

Procedure

First, participants completed a short demographic questionnaire that recorded age, gender and years of education, confirmed the absence of any pre-existing

neurological condition or learning difficulty, and verified that they had normal/corrected-to-normal vision and hearing. Next, tests were administered in the following order: NART17, STW-2, WAIS-IV, and then the remaining NART words. Full NART, NART17 and STW-2 results were used to predict WAIS-IV FSIQ using published regression equations (Bright et al., 2018; van der Linde & Bright, 2018) or manual tables (Baddeley & Crawford, 2012). Participants were informed at the beginning how long each test was likely to take, and reminded that they may take breaks whenever needed. Overall, participation lasted from 90 to 120 minutes. Since the NART is essentially effortless, not requiring prolonged concentration or motivation (Nelson, 1982), we do not envisage that any confound will have arisen from always testing the earlier-identified subset (NART17) before the remaining 33 items. For convenience, the NART17 words and regression equation are reproduced in Appendix.

Results are compared to two earlier attempts to abbreviate the NART. The first is the Mini-NART (McGrory et al., 2015), which uses a subset of 23 words identified through Mokken scaling and provides a regression equation for the calculation of FSIQ. The second is the Short NART (Beardsall & Brayne, 1990), which uses the first half of the NART (25 words) only, and is suggested for use only with those with low reading ability (in practice, those with five or fewer errors on the first half of the test progress to complete the full NART). Here, since the 17-word subset was presented prior to the remaining 33 words, the last 25 words with respect to the original test ordering were removed post-hoc for those participants that qualified for a Short NART score. To calculate FSIQ, the authors provide a table to convert the Short NART score into a full NART score only for those participants that did not go on to complete the full NART, and then refer readers to the NART manual (Nelson, 1982). However, since the NART has recently been restandardized against the most recent revision of the WAIS, and since WAIS-IV was the IQ battery that was used in this study, the updated regression equation in Bright et al. (2018) was used.

GAs simulate Darwinian evolution by survival of the fittest by generating multiple slightly different (mutated) solutions to a problem, typically from a random starting point, and select the best solution(s) generated to seed the subsequent generation, thereby moving closer to an optimal solution over a number of generations (for a more detailed explanation, see Mitchell, 1996). The GA described in van der Linde and Bright (2018) was applied to a bit matrix containing the 100 individual STW-2 responses for each participant, with 1 indicating a lexical decision error (i.e., selection of a pseudoword rather than a real word). Since STW-2 contains 100 word pairs there are a *little googol* (2^{100}) possible subsets of this test – too many to be evaluated by examining them one at a time,¹ and too many to be truly optimised by any conventional statistical process. The GA navigates this vast search space to identify optimal subsets through simulated evolution/hill-climbing. GA settings were: 128 generations per run, 128 children per generation, 10% mutation rate, and

100 complete repetitions. Exhaustive leave-one-out cross-validation (LOOCV; Hastie et al., 2009), was performed to evaluate potential overfitting using the same settings. Using LOOCV, a separate model is constructed using all subsets of the original data in which one single data point (here, participant) has been left out. Each model is used to predict the dependent variable (here WAIS-IV FSIQ) for the single participant that was not used in the training of that model. The average performance for all left-out participants using the respective models in which they were left out provides a measure of how well the model is likely to generalise to new participants.

Results

The mean measured WAIS-IV FSIQ was 110.67 (SD 12.11). Reading test mean score, score range, mean estimated WAIS-IV FSIQ, root mean square error (RMSE), maximum absolute error from measured WAIS-IV FSIQ, and Pearson product moment correlation between estimated and measured FSIQ are shown in Table 1. Correlations between test scores are shown in Table 2; note that the very high correlation between the short NART and Full-NART scores occurred because they were identical for most participants (i.e., most participants did not qualify for early cessation of the test, see below). Scatterplots showing the relationship between estimated and measured WAIS-IV FSIQ for each word test are shown in Figure 1. These results confirm the validity of the NART17 on a new cohort of participants, and demonstrate that it performs at least as well as the full NART and the STW-2 (with numerically lower RMSE, lower maximum absolute error, and higher correlation with WAIS-IV FSIQ), despite using only 17% of the number of test items in STW-2 and one-third of those in the full NART. To identify whether the size of the correlation coefficients between the measures of premorbid test performance and WAIS-IV FSIQ (presented in Table 1) differed significantly, the Williams (1959) test for

Table 1. Number of test items, reading test mean scores, score range, mean estimated WAIS-IV FSIQ, RMSE between estimated and measured WAIS-IV FSIQ, maximum absolute error between estimated and measured WAIS-IV FSIQ, and correlation between estimated and measured WAIS-IV FSIQ.

	STW-2 raw score	Full NART errors	NART17 errors	Short-NART errors	Mini-NART errors
Test items	100	50	17	25 or 50	23
Mean score (SD)	72.82 (11.93)	19.42 (8.10)	7.34 (2.48)	19.87 (8.42)	9.01 (4.04)
Score range	51.97	3.36	2.15	3.38	1.16
Mean estimated FSIQ (SD)	104.90 (7.35)	107.43 (7.92)	107.08 (8.67)	106.99 (8.23)	97.74 (9.47)
Root mean square error (RMSE) from measured FSIQ	11.09	9.79	9.51	9.93	16.22
Max absolute error (MAE) from measured FSIQ	29.00	29.94	22.73	31.89	41.96
Pearson correlation of score with measured FSIQ	0.62	-0.64	-0.69	-0.64	-0.61

Table 2. Absolute correlations between reading test scores.

	STW-2 raw score	Full NART errors	NART17 errors	Short-NART errors	Mini-NART errors
STW-2 Raw Score		0.83	0.81	0.81	0.82
Full NART Errors			0.91	0.98	0.95
NART17 Errors				0.89	0.85
Short-NART Errors					0.94

differences between dependent correlations was applied to provide t-statistics for each comparison. These results show that none of the correlations differed significantly from one another ($p > .1$ in all cases).

For the Short NART, of our 69 participants only 25 made more than 5 errors in the first half of the test, so the full NART error score was required for the remaining 44 participants (nearly two-thirds). For those participants identified as suitable for the Short NART, their full NART score was estimated using the table

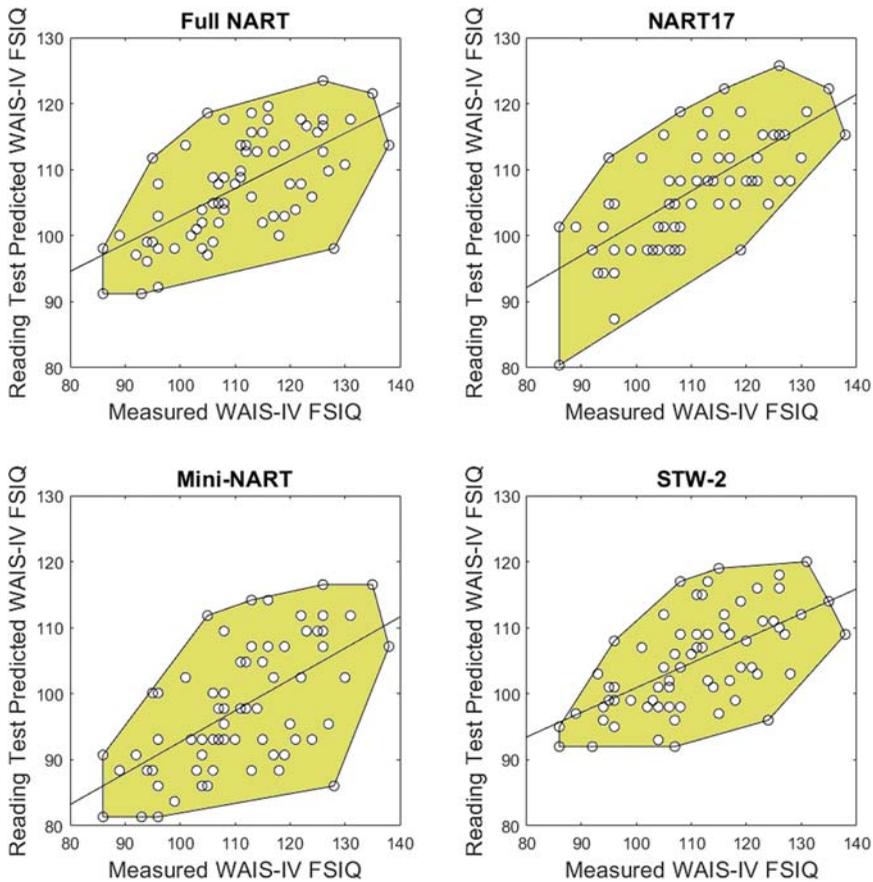


Figure 1. Scatterplots showing estimated and measured WAIS-IV FSIQ for each word test. Black circles correspond to participants. The black line is the line of best fit (least squares) between measured and estimated WAIS-IV FSIQ. The proximity of the participant data to the line of best fit is highlighted as a shaded zone (convex hull).

provided in Beardsall and Brayne (1990), and their full-scale-IQ was then calculated using the regression equation in Bright et al. (2018).

Running the GA on the bit matrix containing participants' STW-2 correct responses (0) and errors (1) reveals a similar degree of redundancy to that detected in our previous work that examined the NART. Over 100 repetitions, the mean test subset size was 35.37 items (SD 2.22), approximately two thirds less than the full 100-item test, with a mean RMSE from the measured WAIS-IV FSIQ of 6.89 (SD 0.05). The mean product moment correlation between the identified subset raw scores and WAIS-IV FSIQ was $r = 0.82$ (SD 0.00). The best test subset, with $r = 0.83$ and $RMSE = 6.76$, included the following 34 test item numbers: (1,2,3,5,6,7,10,11,19,23,26,27,28,32,35,44,50,52,53,56,57,59,62,63,66,73,78,82,83,87,88,90,91,95) [note that the word stimuli associated with each test item are not reproduced here to protect copyright]. A direct statistical comparison with the full STW-2/WAIS-IV FSIQ correlation ($r = .83$ vs $r = .62$) using Williams (1959) method shows that the 34-item subset provided a significant improvement over the full STW-2 in predictive accuracy [$t(66) = 6.36$, $p < .001$]. To use this subset, the regression equation for predicting WAIS-IV FSIQ (\hat{y}) from the 34-item STW-2 subset score (x) would be $\hat{y} = ax + b$ with constants $a = -2.7571$ and $b = 131.5247$. Although the RMSE and correlation above show statistically superior performance to the measures in Table 1, it is important to note that the GA is susceptible to overfitting the model; that is, basing the subset identified too closely to the specific data used to train it. For this reason, cross-validation is necessary.

LOOCV is exhaustive, holding back each participant, in turn, training the model on the remaining $N-1$ participants, and then testing the accuracy of each of the N models produced on one participant not used to train that model (i.e., left out). Average prediction error for the left-out participants is the critical statistic, to be measured as RMSE. The RMSE of the N models on the corresponding left out participants was 10.40 (SD 11.51). Although the RMSE prediction for the N left-out participants on their corresponding models was slightly higher than the models that had been trained using data from all participants (mean 6.89, best 6.76, as noted above), this is to be expected as the validation models were asked to make a prediction on data that they had not been explicitly trained on. Furthermore, importantly, the RMSE between STW-2 predicted WAIS-IV FSIQ using the instructions in the official manual and measured WAIS-IV FSIQ was still higher, at 11.09, adding weight to the argument that two-thirds of the STW-2 test items are ostensibly redundant (i.e., they add no additional accuracy to the WAIS-IV FSIQ estimate produced).

Discussion

This article incorporates three main findings. First, using 69 new participants, we tested an earlier model trained using $n = 92$ participants in which the NART

premorbid IQ test was shortened by two-thirds. Our findings confirm that the shorter test does not reduce WAIS-IV FSIQ prediction accuracy. Furthermore, the genetic algorithm-derived test (NART17) was found to perform at least as well as two earlier attempts to abbreviate this test. The 23-word Mini-NART proposed using Mokken scaling (McGrory et al., 2015) produced significantly larger RMSE in WAIS-IV FSIQ prediction, the largest single prediction error, and the numerically smallest absolute correlation. For the Short NART, only 25 of our 69 participants were eligible (around one-third), and RMSE was fractionally smaller than that achieved using the NART17. The NART17 also had the smallest maximum absolute prediction error, and the numerically highest correlation with measured FSIQ. Furthermore, since the NART17 is a subset of the full NART, studies showing high inter-rater reliability (Crawford, 1992; O'Carroll, 1987) and test-retest reliability (Crawford, 1992) still apply.

Second, the accuracy of NART and STW-2 in predicting FSIQ were compared and it was found that both full NART and NART17 offer statistically equivalent precision (we note here that, since STW-2 does not require speech, it remains particularly useful in scenarios in which speech may have been affected by a neurological condition, or in which speech production would elicit stress).

Third, we computed that, remarkably, two-thirds of the 100 word pairs in the STW-2 test can be eliminated with no corresponding decrease in WAIS-IV FSIQ prediction accuracy. While even modest reductions in assessment time may be advantageous for clinicians and their patients, our validation of these abbreviated tools for estimating premorbid FSIQ serves to demonstrate the power of the method proposed in van der Linde and Bright (2018). Just as we confirmed/validated NART17 in this article using a new participant group, it would also be prudent to: (i) validate the abbreviated STW-2 findings with new participants in future work, and; (ii) to examine other dominant tests of premorbid ability that use word reading, such as the 70-word Test of Premorbid Functioning (TOPF; Wechsler, 2011).

In summary, our data indicate that full NART and NART17 predict FSIQ at least as well as the full STW-2. With validity examined through exhaustive cross-validation, the GA also determined that only around one-third of the 100 words pairs in STW-2 were necessary to confer very high predictive accuracy, suggesting that consideration should also be given to this test being abbreviated to support efficient clinical assessment. Moreover, in comparing the performance of NART17 to two previous attempts to shorten this test, the Mini-NART (McGrory et al., 2015) and the Short NART (Beardsall & Brayne, 1990), we found the NART17 to produce numerically more precise (especially compared to Mini-NART) and statistically equivalent estimates, and to be usable in a wider range of circumstances than the Short NART.

Our earlier work indicates that tests based on reading ability and vocabulary knowledge offer valid estimates of premorbid ability in a range of neurological conditions (Bright et al., 2002). Nevertheless, scores on such tests are

likely to deteriorate in conditions such as semantic dementia, moderate and/or late Alzheimer's disease, and in other disorders associated with a surface dyslexia. We encourage further work aimed at determining test validity in these patients.

The validation of the NART17 demonstrates that GA-based approaches are both effective in identifying optimum test question subsets from existing tests, and also have the potential to be used in test development by identifying a core subset of words or questions from a larger corpus that confer the greatest estimation accuracy for a measure of interest, rather than relying upon qualitative approaches like those used in the original development of the NART.

Note

1. If we assume that each subset takes 1 ms to evaluate, evaluating 2^{100} subsets would take 10^{30} ms, which is equal to 10^{19} years.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ian van der Linde  <http://orcid.org/0000-0002-8131-5906>

Peter Bright  <http://orcid.org/0000-0003-1186-8035>

References

- Baddeley, A., & Crawford, J. (2012). *Spot the word*. Pearson Assessment.
- Beardsall, L., & Brayne, C. (1990). Estimation of verbal intelligence in an elderly community: A prediction analysis using a shortened NART. *British Journal of Clinical Psychology*, 29(1), 83–90. <https://doi.org/10.1111/j.2044-8260.1990.tb00851.x>
- Bright, P., Hale, E., Gooch, V. J., Myhill, T., & van der Linde, I. (2018). The National Adult Reading Test: Restandardisation against the Wechsler Adult Intelligence Scale – Fourth edition. *Neuropsychological Rehabilitation*, 28(6), 1019–1027. <https://doi.org/10.1080/09602011.2016.1231121>
- Bright, P., Jaldow, E., & Kopelman, M. D. (2002). The National Adult Reading Test as a measure of premorbid intelligence: A comparison with estimates derived from demographic variables. *Journal of the International Neuropsychological Society*, 8(6), 847–854. <https://doi.org/10.1017/S1355617702860131>
- Bright, P., & van der Linde, I. (2020). Comparison of methods for estimating premorbid intelligence. *Neuropsychological Rehabilitation*, 30(1), 1–14. <https://doi.org/10.1080/09602011.2018.1445650>
- Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, D. M. Parker, & W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21–49). Erlbaum.

- Crawford, J. R., O'Carroll, R. E., & Venneri, A. (1998). Neuropsychological assessment of the elderly. In A. S. Bellack, & M. Hersen (Eds.), *Comprehensive clinical psychology, volume 7: Clinical geropsychology* (pp. 133–169). Pergamon.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction*. Springer-Verlag.
- McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., Whalley, L. J., & Deary, I. J. (2004). Pronunciation of irregular words is preserved in dementia, validating pre-morbid IQ estimation. *Neurology*, *62*(7), 1184–1186. <https://doi.org/10.1212/01.WNL.0000103169.80910.8B>
- McGrory, S., Austin, E. J., Shenkin, S. D., Starr, J. M., & Deary, I. J. (2015). From “aisle” to “labile”: A hierarchical National Adult Reading Test scale revealed by Mokken scaling. *Psychological Assessment*, *27*(3), 932. <https://doi.org/10.1037/pas0000091>
- Mitchell, M. (1996). *An introduction to genetic algorithms*. MIT Press.
- Nelson, H. E. (1982). *The National Adult Reading Test (NART)*. NFER-Nelson.
- Nelson, H. E., & Willison, J. (1992). *The National Adult Reading Test (NART)*. NFER-Nelson.
- O'Carroll, R. (1995). The assessment of premorbid ability: A critical review. *Neurocase*, *1*(1), 83–89. <https://doi.org/10.1080/13554799508402350>
- 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*(3), 229–230. <https://doi.org/10.1111/j.2044-8260.1987.tb01352.x>
- Sharpe, K., & O'Carroll, R. (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: A Canadian study. *British Journal of Clinical Psychology*, *30*(4), 381–384. <https://doi.org/10.1111/j.2044-8260.1991.tb00962.x>
- van der Linde, I., & Bright, P. (2018). A genetic algorithm to find optimal reading test word subsets for estimating full-scale IQ. *PLoS One*, *13*(10), e0205754. <https://doi.org/10.1371/journal.pone.0205754>
- Wechsler, D. (2008). *Wechsler adult intelligence scale*. Pearson Assessment.
- Wechsler, D. (2011). *Test of premorbid functioning. UK version (TOPF UK)*. Pearson Corporation.
- Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *21*(2), 396–399. <https://doi.org/10.1111/j.2517-6161.1959.tb00346.x>

Appendix. NART17 test words and regression equation

1: Capon	7: Gouge	13: Puerperal
2: Nausea	8: Superfluous	14: Leviathan
3: Courteous	9: Simile	15: Prelate
4: Naïve	10: Façade	16: Demesne
5: Thyme	11: Drachm	17: Campanile
6: Procreate	12: Idyll	

WAIS-IV Full Scale IQ Prediction = $(-3.4882 \times \text{NART17_Errors}) + 132.7113$.