



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936

Citation for published version:

Dykiert, D & Deary, IJ 2013, 'Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936', *Psychological assessment*, vol. 25, no. 4, pp. 1361-1366. <https://doi.org/10.1037/a0033623>

Digital Object Identifier (DOI):

[10.1037/a0033623](https://doi.org/10.1037/a0033623)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Psychological assessment

Publisher Rights Statement:

© Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936. / Dykiert, Dominika; Deary, Ian J.
In: Psychological assessment, Vol. 25, No. 4, 12.2013, p. 1361-1366.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Running Head: Retrospective validation of WTAR and NART

Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936

Dominika Dykiert and Ian J. Deary

Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology,
University of Edinburgh

Author Note

Dominika Dykiert, Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh; Ian J. Deary, Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh.

Acknowledgements. Lothian Birth Cohort 1936 study has been supported by: Research into Ageing (Wave 1), Age UK and the Medical Research Council (Waves 2 and 3). The work was undertaken at The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

Correspondence concerning this article should be addressed to Dominika Dykiert, Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland, UK. E-mail: ddykiert@staffmail.ed.ac.uk

Abstract

In order to assess the degree of cognitive decline resulting from a pathological state, such as dementia, or from a normal ageing process, it is necessary to know or to have a valid estimate of premorbid (or prior) cognitive ability. National Adult Reading Test (NART) and Wechsler Test of Adult Reading (WTAR) are two tests developed to estimate premorbid or prior ability. Due to the rarity of actual prior ability data, validation studies usually compare NART/WTAR performance with measures of current abilities in pathological and non-pathological groups. In this study, we validate the use of WTAR scores and extend the validation of the use of NART scores as estimates of prior ability, vis-à-vis the actual prior (childhood) cognitive ability. We do this in a large sample of healthy older people, the Lothian Birth Cohort of 1936. Both NART and WTAR scores were correlated with cognitive ability tested in childhood ($r = .66-.68$). Scores on both the NART and WTAR had high stability over a period of 3 years in old age (r in excess of .90), and high inter-rater reliability. NART accounted for more unique variance in childhood intelligence than did WTAR.

Key words: NART, WTAR, premorbid cognitive ability

Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936

Introduction

Estimating a person's level of premorbid cognitive ability is important in both clinical and research settings. From a clinician's perspective, knowing a patient's current level of cognitive ability is not, in itself, necessarily informative in terms of the amount of decrement, for example resulting from a traumatic brain injury (TBI), due to a neurodegenerative disorder, or the ageing process. The same low current cognitive ability score may be obtained by an individual who used to be of average or superior ability and who has suffered decline, or by an unimpaired individual whose prior level of ability was low. In order to draw conclusions about change, it is necessary to know or to have a valid estimate of premorbid (before trauma or the onset of illness) or prior (before the effects of aging) ability; that is, the level of ability before the onset of decline from whatever cause(s).

One approach to this exploits the distinction between fluid and crystallised cognitive abilities (Horn & Cattell, 1967). Fluid abilities refer to basic information processing and are involved in logical reasoning, inducing general rules from patterns and analogies, and applying and extrapolating them to solve novel problems. Crystallised abilities, on the other hand, refer to acquired knowledge—they are learned through experience and stored. Crystallised abilities are sometimes preserved in cases in which fluid abilities have been affected. For example, reading ability is preserved in patients diagnosed with dementia (Nelson & McKenna, 1975). This discovery brought the idea of a test to estimate premorbid ability. The attraction of such test was that it could be administered to affected individuals, for example following TBI or those suffering from dementia, to obtain an estimate of their cognitive performance before it was affected.

This was put into practice by Nelson and O'Connell (1978), who developed NART – the New (later, National) Adult Reading Test. The NART involves the participant reading aloud a list of 50 words with irregular grapheme-phoneme mappings or stress patterns (e.g. *ache*, *naive*). The person's test score is the number of items in which pronunciation errors are made. The main premise behind this test is that the words can be read and pronounced correctly only if they are familiar to the participant; relying on normal pronunciation rules would result in an error. Therefore, participants need to rely on their prior knowledge of the word and of its pronunciation, which remains relatively unimpaired even when other abilities have been affected, for example in dementia (Nelson & McKenna, 1975). A more recently-developed test, formally similar to NART, but using a different set of words, is the Wechsler Test of Adult Reading (WTAR, The Psychological Corporation, 2001). It was specifically developed and co-normed with Wechsler Adult Intelligence Scale–III (Wechsler, 1997).

Validation of premorbid ability estimates is problematic for the same reason that they needed to be developed—the scarcity of actual premorbid cognitive ability data. The usual type of study to establish the validity of NART or WTAR scores as estimators of premorbid ability in clinical populations is performed using a case-control paradigm. In a typical study, a group of affected individuals is compared with a group of controls on a putative premorbid ability estimation test and on a test or tests assessing current (fluid) cognitive ability. The expectation is that, if a test indexes premorbid rather than current ability then, provided the groups are otherwise well matched: a) there will be a difference between groups in current ability but not in premorbid ability; and/or b) the estimated premorbid IQ will be higher than the current IQ in the affected group but not in the control group, or the decrement between premorbid and current cognitive ability will be greater in the patient group compared with the control group. The use of NART and WTAR scores as estimates of premorbid ability has been validated in TBI (Green et al., 2008; Moss & Dowd, 1991; Watt & O'Carroll, 1999),

mild dementia (Crawford, Parker & Besson, 1988; Nelson & O'Connell, 1978; McGurn et al, 2004), and other neurological disorders (see Nelson & Willison, 1991; The Psychological Corporation, 2001, for reviews). For example, Watt and O'Carroll (1999) compared NART performance of 25 traumatic brain injury patients, 20 orthopaedic patients with no brain injury, and 50 healthy individuals. After controlling for education, sex and depression, the groups had comparable NART performance, whereas there were significant differences in current ability between the groups—TBI patients scored lower than orthopaedic or healthy controls. Cognitive ability declines not only as a result of a clinical condition but also in normal cognitive ageing. There are individual differences in the rate of cognitive decline with age even within the normal range (Wilson et al., 2002). Cognitive decline may lead to loss of independence and lower wellbeing; it is a burden to those affected, their families and the wider society (Deary et al., 2009). It is important to distinguish between those with greater and lesser amounts of cognitive ageing in order to identify those at a higher risk of negative outcomes, such as dementia, or even death, and to study the determinants of the individual differences in cognitive ageing. NART and WTAR may be useful in both of the above cases.

To date, few studies attempted to validate the use of NART scores as estimates of prior ability in healthy samples of older people; we know of no studies validating the use of WTAR scores in this context. Berry et al. (1994) reported a high correlation ($r = .70$) between WAIS-R full-scale IQ (mean age 67.8 years) and WAIS-R IQ estimated from a North American version of NART (Blair & Spreen, 1989) obtained 3.5 years later. One limitation of this study was that the WAIS-R was administered relatively close in time to the NART. The gap of 3.5 years may not be sufficient to incur much ageing effect, and the age at which the WAIS-R was administered will mean that there will have been ageing-related decrements on the actual prior ability. Therefore, the high correlation between WAIS- and NART-estimated IQs might reflect sensitivity of both these measures to cognitive decline.

Studies having truly prior cognitive ability and then cognitive ability tested in old age are rare. Crawford, Deary, Starr, and Whalley (2001) were able to test the association between NART administered at age 77 and a general cognitive ability test (an adaptation of the Moray House Test No. 12, MHT) administered 66 years earlier. They found a correlation of .73. Because the prior ability was measured at age 11 years, it is safe to assume that it was not affected by age-related cognitive decline.

Although several studies have attempted to validate the use of NART scores as a measure of prior or premorbid ability, WTAR has received much less research attention. Therefore the present study set out, in a large sample of healthy older people, to extend validation testing of the use of NART scores and to test the validity of the use of WTAR scores as estimates of prior cognitive ability. The sample is the Lothian Birth Cohort 1936 (Deary et al., 2007; Deary, Gow, Pattie & Starr, 2011), who undertook a test of cognitive ability in childhood and again in later life. Therefore, a true measure of prior ability is available for this group, as well as measurements on the same cognitive test taken six decades later.

The aims of this study were to test validity of using NART and WTAR scores as estimates of prior cognitive ability in a non-clinical sample of older people; and to compare associations of NART and WTAR with prior (childhood) and current (old-age) ability.

Method

The participants were members of the Lothian Birth Cohort 1936 (LBC1936), a longitudinal study focussed on cognitive ageing. The recruitment and testing of the participants is described in two open-access reports (Deary et al., 2007; Deary et al., 2011) and is given here briefly.

Almost all children born in 1936 and attending schools in Scotland in 1947 (N=70,805) took, at a mean age of 11 years, a general intelligence test as part of the Scottish

Mental Survey 1947 (Scottish Council for Research in Education [SCRE], 1949). Surviving members of the survey, who lived in Edinburgh (Scotland) or the surrounding area, were invited to take part in the LBC1936 study. They were identified from a list of people registered with a general medical practitioner, which captures most of the Scottish population. Invitations were sent out to 3,686 potentially eligible participants, of whom 1,703 responded (46%). Those who were eligible and who agreed to take part in the study were followed up from 2004 onwards. So far, there have been three waves of testing. Wave 1 took place in 2004-2007 and, in it, 1,091 participants (548 men, 543 women; 64% of those who responded to the invitation) were tested at a mean age of 69.5 years ($SD = 0.8$). Wave 2 took place in 2007-2010 and 866 participants (448 men, 418 women) were tested at a mean age of 72.5 ($SD = 0.7$). Wave 3 is being conducted at the time of writing. At each wave of testing, the participants complete a battery of cognitive tests, undergo physical examinations, and complete self-report questionnaires. The tests used in the present study are described next.

The participants took a cognitive test at age 11 and completed a battery of tests aged approximately 70 and 73 years (waves 1 and 2). All tests were administered and scored by trained psychologists. There were 8 testers overall, with 5 testers involved in cognitive testing at wave 1, and 5 at wave 2.

The tests measuring current fluid cognitive ability used in this study was a modified version of the Moray House Test (MHT) No. 12, (SCRE, 1933). MHT is a paper-and-pencil test of general intelligence, which was administered to the participants at age 11 and again at age 70. It has a maximum score of 76 and comprises items of different types, including: following instructions, same-opposites, odd-one-out word classifications, analogies, practical, reasoning, proverbs, arithmetical, geometrical, jumbled sentences, cypher decoding, and 4 other, unclassified, items. MHT scores in childhood have high correlations with the Stanford-Binet intelligence test, $r = .78$ in girls ($N = 500$) and $r = .81$ in boys ($N = 500$) (SCRE, 1933).

Scores on the MHT obtained from 70-year-olds also have a high correlation ($r = .67$) with concurrently-obtained scores on a general cognitive factor obtained from six non-verbal subtests of the Wechsler Adult Intelligence Scale-III^{UK} (Deary, Johnson, & Starr, 2010). Scores on MHT obtained from 80-year-olds also have high a correlation with concurrently-obtained scores on Raven's Standard Progressive Matrices ($r = .71$ for men and women; Deary, Whiteman et al., 2004). MHT was to be completed within a time limit of 45 minutes. The same instructions were used in childhood and at age 70.

Prior ability was estimated using two reading tests: NART (Nelson & Willison, 1991) and WTAR (The Psychological Corporation, 2001). Both were administered at waves 1 and 2 of the LBC1936 study. Both NART and WTAR comprise 50 short, irregular words, which the participants are asked to read aloud. The words are arranged in order of difficulty, from relatively easy (e.g. NART: chord, WTAR: again) to more difficult (e.g. NART: campanile, WTAR: hegemony). To make the results as clear as possible, we used the number of correct responses from NART and WTAR as their outcomes, so that higher scores indicate better performance. For some additional analyses, we predicted full scale IQs based on the NART or WTAR (Nelson & Willison, 1991; The Psychological Corporation, 2001 [UK norms]).

As reported in their respective manuals (Nelson & Willison, 1991; The Psychological Corporation, 2001), NART and WTAR scores have high split-half (NART: .93, WTAR: .87-.95) and high test-retest reliabilities (NART: .98, WTAR: .90-.94). The interpretation of NART and WTAR scores as estimators of general cognitive ability in normal populations has also been shown to be valid.

For the present study, only participants who attended waves 1 and 2 of testing and had complete data on the MHT, NART and WTAR were included. One participant was excluded because of an outlying score on the MHT at age 11. Except where raw scores are used for summaries, NART, WTAR, and MHT scores were adjusted for age in days at testing.

Participants with Mini Mental State Examination (MMSE; Folstein, Folstein & McHugh 1975) score of below 24 at either wave were also excluded ($n = 16$), as such low scores indicate possible dementia. The working sample included 802 participants (413 males, 389 females).

Results and Discussion

Descriptive statistics for the main variables of interest are presented in Table 1. Raw WTAR scores were, on average, higher than raw NART scores obtained at the same wave (both tests had 50 items). At wave 1 raw WTAR scores were higher than NART scores by 6.56 points, $t = 51.59$, $df = 801$, $p < .001$, a difference very similar in magnitude to one reported by McFarlane, Welch, and Rodgers (2006). More people scored the maximum on WTAR ($n=46$) than on the NART ($n=1$). The mean NART-estimated IQ (112.06, $sd = 9.74$) was higher than mean WTAR-estimated IQ (107.03, $sd = 7.93$), $t = 31.30$, $df = 801$, $p < .001$. At wave 1 correlation between NART and WTAR was high and positive, $r = .89$, $p < .001$ whether calculated using raw scores or predicted IQs. This correlation was similar to the one previously-reported in a neurologically unimpaired sample ($r=.84$, Mathias, Bowden, Barrett-Woodbridge, 2007), and higher than the correlation found in clinical samples (e.g. $r=.78$; The Psychological Corporation, 2001). Raw MHT scores obtained in old age were higher than those obtained in childhood, $t = -50.63$, $df = 801$, $p < .001$.

Small decreases in performance on the reading tests were observed over the 3-year period from age 70 to age 73. Mean NART score decreased by 0.43 ($d = 0.05$), $t = 4.29$, $df = 801$, $p < .001$. Mean WTAR score decreased by 0.34 ($d = 0.05$), $t = 3.74$, $df = 801$, $p < .001$. The correlations between scores obtained at wave 1 and wave 2 were similarly high for both tests, indicating test-score stability over time: $r = .94$, $p < .001$ for NART; and $r = .93$, $p < .001$ for WTAR. The magnitude of the coefficients is comparable with those reported previously, over shorter periods of time. For example, Crawford, Parker, Stewart, Besson, and

De Lacey (1989) reported test-retest reliability of the NART scores over 10 days to be .98. Deary, Whalley & Crawford, (2004) reported a correlation of .89 between NART scores obtained 1 year apart. Test-retest stability of WTAR over 2 to 12 weeks in a sample of 55 to 74-year-olds was estimated to be .94 (The Psychological Corporation, 2001).

As well as indicating stability of NART and WTAR scores over the 3-year period, the high correlations between wave 1 and wave 2 indicate high consistency of testers. However, the correlations do not indicate the level of agreement between the testers. NART scores have been reported to have high inter-rater reliability before, but this view was based largely on the reports of high correlations between different raters of the pronunciation of NART words (e.g. O'Carroll, 1987). Here, we tested inter-rater reliability by a series of ANOVAs with testers as fixed factors and by calculating intra-class correlations on a subsample of our testers. Details of these analyses can be found in the Supplement. In brief, our study adds evidence that different raters not only rank participants in the same order, but they also agree on the absolute score for each individual. The latter cannot be determined by a between-tester correlation.

The pattern of correlations between NART and WTAR and MHT IQ measured at age 11 and age 70 are shown in Figure 1. NART was similarly associated with MHT IQ at age 11 as it was with MHT IQ at age 70 ($r = .68$, and $.67$, respectively; $p < .001$, for both). We used Williams's Test (1959), implemented in the statistical software R (paired.r command) to test whether the correlations were significantly different. They were not ($t=0.52$, $df=799$, $p=.602$). The correlation between MHT IQ measured at age 11 and at age 70 was $r = .69$, $p < .001$. This correlation was not significantly different from the correlation between age 11 MHT IQ and NART ($t = -0.52$, $df = 799$, $p = .605$). In other words, childhood MHT IQ was as highly correlated with scores obtained on NART as it was with MHT IQ measured in later life.

The associations of WTAR with age MHT 11 IQ and age 70 MHT IQ were very similar: $r = .66$, $p < .001$ and $r = .65$, $p < .001$, respectively. They were not statistically different from each other, $t = -0.51$, $df = 799$, $p = .613$. Similarly to NART, WTAR was as strongly associated with age 11 MHT IQ as age 11 MHT IQ was with age 70 IQ, $t = -1.49$, $df = 799$, $p < .137$. When instead of raw WTAR scores we used IQ estimated from WTAR score, augmented by demographic variables (The Psychological Corporation, 2001), correlations with age 11 IQ and with age 70 IQ (for both, $r = .65$, $p < .001$) were almost unchanged, and not significantly different.

NART and WTAR were equally strongly associated with age 11 MHT IQ ($t = 1.66$, $df = 799$, $p = .096$) and age 70 MHT IQ ($t = 1.64$, $df = 799$, $p = .101$). Given the strong correlation between NART and WTAR, it was desirable to partition common and unique variance of each of them in (retrospectively) predicting prior ability. To do this, we performed commonality analysis (Nimon, Lewis, Kane & Haynes, 2008). Taken together, NART and WTAR accounted for 46% of variance in age 11 MHT IQ. As expected, there was a large overlap between NART and WTAR, with 88% of the explained variance shared between them. NART contributed 10% of unique variance and WTAR contributed a further 2%.

In an earlier study of NART, Crawford et al. (2001) argued that if it is primarily a measure of prior ability, it should correlate at least as highly with childhood ability as with current ability in old age. Conversely, NART's higher correlation with current than with prior ability would suggest it is primarily a measure of current level of cognitive functioning. Crawford et al. found that NART correlated equally strongly with cognitive ability measured in childhood and in old age. Following the above-mentioned reasoning, we confirmed in a much larger study that NART meets the criterion for measures of prior cognitive ability. We also demonstrated, for the first time, that WTAR also meets this criterion.

In our study NART explained slightly more variance in childhood IQ than WTAR. Earlier evidence showed that NART predicted concurrent level of cognitive ability slightly better than WTAR (Mathias et al., 2007). Therefore, although the two tests are formally similar and scores on them are highly correlated, there is perhaps a slight advantage for the NART.

The study had a number of strengths. First, it had the rarely-available measure of actual prior ability test from six decades before the tests in old age. The use of MHT scores as measures of general cognitive ability in childhood (SCRE, 1933; 1949) and old age (Deary et al., 2010; Deary, Whiteman, et al., 2004) are well-validated. The administration of MHT at age 11 ensured that it is free from age-related decline. Therefore, we were in a rare position retrospectively to validate the use of NART and WTAR scores as estimates of prior ability. However, the childhood MHT score is likely to be an underestimate of the maximal ability, as it was obtained before the individuals reached their peak of crystallised intelligence in adulthood. It is expected that, with the growth that followed, there occurred some re-assortment of individual differences. In other words, the age-11 estimate of the rank order in terms of peak ability is less than perfect. The extent of this association would act to reduce the observed effects. Additional strengths are the large size and age-homogeneity of the LBC1936 sample.

One limitation of the present study is that the LBC1936 is not fully representative of the population. Surviving members of the cohort, who agreed to take part in the study aged 70 and 73 have, on average, higher cognitive abilities than those who did not participate. Full scale IQs predicted from mean NART and WTAR score at wave 1 were both above average. The restriction of range in the sample probably reduced the effect sizes below the true effects. Another related limitation is that having a relatively high cognitive ability sample meant that some of the distributions were negatively skewed, with a proportion of participants

approaching or reaching the ceiling. This was more evident for WTAR than NART and more for MHT scores at age 70 than at age 11. Performing a similar study with tests that are more sensitive at the upper ranges of cognitive ability would be a useful extension to this investigation. Finally, both the reading tests and the MHT are verbal. Knowing that NART and WTAR scores are more strongly correlated with verbal than non-verbal intellectual abilities (Crawford et al., 1989; The Psychological Corporation, 2001), effect sizes might have been smaller if we had used a multi-domain test of intelligence. However, in mitigation of this, we also note that: the MHT principally tests reasoning with simple words and numbers, and mostly does not test vocabulary per se; and the MHT was applied at an age before much of the vocabulary tested by the NART and WTAR were developed.

In the present sample we used an older-age UK (Scottish) sample. WTAR has been standardised for both the UK and US. NART has been developed and standardised in the UK, and versions of the test have been developed, specifically for use with North American samples (see Blair & Spreen, 1989; Grober & Sliwinski, 1991).

Our findings show that NART and WTAR can be used as estimators of prior ability in a healthy sample of older people, in that are strongly related to IQ measured in childhood. However, it should be noted that previous studies suggested that, if NART and WTAR are used to estimate IQ, there might be variations in the accuracy of such predictions depending on the level of cognitive ability. Specifically, both NART and WTAR apparently overestimate the IQs of people with low abilities and underestimate the IQs of people with high abilities (e.g. Mathias et al., 2007). Therefore, care needs to be taken when using the tests with samples with relatively high or low cognitive ability. Moreover, NART and WTAR may underestimate premorbid ability of individuals with severe deficits in cognitive functioning (e.g. McFarlane et al., 2006).

In conclusion, we have validated the use of scores obtained on the NART and, for the first time, the WTAR scores in estimating prior cognitive ability in a non-clinical sample of older adults. NART and WTAR behave very similarly in situations considered in this study. Scores on both the NART and WTAR show high stability over time and high inter-rater reliability.

References

- Berry, D. T., Carpenter, G. S., Campbell, D. A., Schmitt, F. A., Helton, K., & Lipke-Molby, T. (1994). The New Adult Reading Test-Revised: accuracy in estimating WAIS-R IQ scores obtained 3.5 years earlier from normal older persons. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, 9, 239–250.
- Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the national adult reading test. *Clinical Neuropsychologist*, 3, 129–136.
doi:10.1080/13854048908403285
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crawford, J. R., Deary, I. J., Starr, J., & Whalley, L. J. (2001). The NART as an index of prior intellectual functioning: a retrospective validity study covering a 66-year interval. *Psychological Medicine*, 31, 451–458. doi:10.1017/S0033291701003634
- Crawford, J. R., Parker, D. M., & Besson, J. A. (1988). Estimation of premorbid intelligence in organic conditions. *The British Journal of Psychiatry: The Journal of Mental Science*, 153, 178–181.
- Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A. O., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension, Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology*, 28, 267–273.
doi:10.1111/j.2044-8260.1989.tb01376.x
- Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., ..., Starr, J.M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, 92, 135–152.
doi:10.1093/bmb/ldp033

- Deary, I. J., Gow, A. J., Pattie, A., & Starr, J. M. (2011). Cohort Profile: The Lothian Birth Cohorts of 1921 and 1936. *International Journal of Epidemiology*. doi:10.1093/ije/dyr197
- Deary, I. J., Gow, A. J., Taylor, M. D., Corley, J., Brett, C., Wilson, V., ..., Starr, J.M. (2007). The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatrics*, 7, 28. doi:10.1186/1471-2318-7-28
- Deary, I. J., Johnson, W., & Starr, J. M. (2010). Are processing speed tasks biomarkers of cognitive aging? *Psychology and Aging*, 25, 219–228. doi:10.1037/a0017750
- Deary, I. J., Whalley, L. J., & Crawford, J. R. (2004). An “instantaneous” estimate of a lifetime’s cognitive change. *Intelligence*, 32, 113–119.
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, 86, 130–147. doi:10.1037/0022-3514.86.1.130
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Green, R., Melo, B., Christensen, B., Ngo, L., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: an examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology*, 30, 163–72.
- 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.

- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129. doi:10.1016/0001-6918(67)90011-X
- Mathias, J.L., Bowden, S.C., & Barrett-Woodbridge, M. (2007). Accuracy of the Wechsler Test of Adult Reading (WTAR) and National Adult Reading Test (NART) when estimating IQ in a healthy Australian sample. *Australian Psychologist*, 42, 49 – 56. doi: 10.1080/00050060600827599
- McFarlane, J., Welch, J., & Rodgers, J. (2006). Severity of Alzheimer’s disease and effect on premorbid measures of intelligence. *Journal of Clinical Psychology*, 453–63. doi: 10.1348/014466505X71245
- McGurn, B., Starr, J., Topfer, J., Pattie, A., Whiteman, M., Lemmon, H., ..., Deary, I.J. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology*, 62, 1184–6.
- Moss, A.R. & Dowd, T. (1991). Does the NART hold after head injury? A case report. *British Journal of Clinical Psychology*, 30, 179-180.
- Nelson, H. E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *The 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., & Willison, J. (1991). *National Adult Reading Test (NART) Test Manual* (2nd ed.). Windsor, UK: NFER-NELSON Publishing Company Ltd.
- Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: an introduction to the package and a practical example. *Behavior Research Methods*, 40, 457–466. doi: 10.3758/BRM.40.2.457

- O'Carroll, R. E. (1987). The inter-rater reliability of the National Adult Reading Test (NART): a pilot study. *The British Journal of Clinical Psychology*, 26, 229–230.
- The Psychological Corporation (2001). *Wechsler Test of Adult Reading*. San Antonio, TX: The Psychological Corporation.
- Scottish Council for Research in Education. (1933). *The Intelligence of Scottish Children: A national survey of an age-group*. London: University of London Press, Ltd.
- Scottish Council for Research in Education. (1949). *The Trend of Scottish Intelligence: A Comparison of the 1947 and 1932 Surveys of the Intelligence of Eleven-year-old Pupils*. London: University of London Press, Ltd.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428.
- Watt, K. J., & O'Carroll, R. E. (1999). Evaluating methods for estimating premorbid intellectual ability in closed head injury. *Journal of Neurology, Neurosurgery & Psychiatry*, 66, 474–479. doi:10.1136/jnnp.66.4.474
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-Third Edition*. San Antonio, TX: The Psychological Corporation.
- Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B*, 21, 396–399.
- Wilson, R. S., Beckett, L. A., Barnes, L. L., Schneider, J. A., Bach, J., Evans, D. A., & Bennett, D. A. (2002). Individual differences in rates of change in cognitive abilities of older persons. *Psychology and Aging*, 17, 179–193.

Table 1. Means, standard deviations and ranges of the cognitive tests from childhood and at 70 and 73 years (N=802)

| | Mean | Std. Deviation | Range |
|----------------------------|-------|-------------------|-----------|
| NART | | | |
| Wave 1 | 35.05 | 7.86 | 12 to 50 |
| Wave 2 | 34.62 | 7.85 | 9 to 50 |
| Change ^a | -0.43 | 2.82 | -12 to 10 |
| WTAR | | | |
| Wave 1 | 41.61 | 6.73 | 16 to 50 |
| Wave 2 | 41.27 | 6.62 | 18 to 50 |
| Change ^a | -0.34 | 2.58 | -11 to 8 |
| MHT 11 | 49.75 | 11.87 | 1 to 74 |
| MHT 70 | 65.26 | 7.75 | 29 to 76 |

Note. NART = National Adult Reading Test, WTAR = Wechsler Test of Adult Reading,

MHT = Moray House Test. Maximum scores were 50 for NART and WTAR and 76 for MHT

^a Negative change indicates decline from wave 1 to wave 2

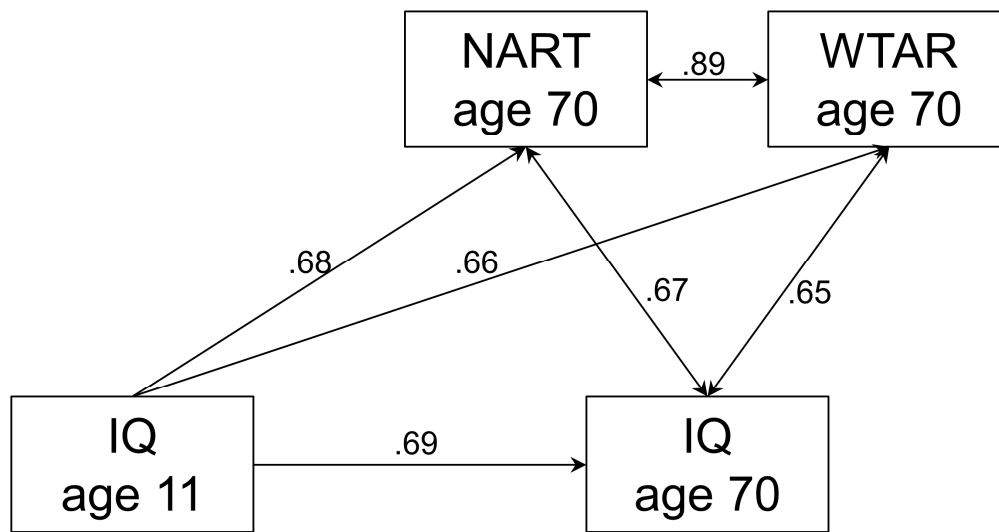


Figure 1. Correlations between IQ at age 11 and NART, WTAR and IQ at age 70

Inter-rater reliability

Calculating inter-rater reliability of NART and WTAR scores was complicated by the fact that there was a 3-year period between the testing sessions. Therefore, estimates of inter-rater reliability will have been lowered because of any instability in NART and WTAR scores over time. However, as reported earlier, the stabilities were very high. Moreover, although there were 8 testers for which inter-rater reliability is of interest, at most two tested a given participant in the two waves considered here. In this case the appropriate measure of inter-rater reliability is a one way model ICC (Shrout & Fleiss, 1979) which treats variance due to raters as error. This is not ideal. Instead, we present means and standard deviations for each tester at both waves (see Table S1).

Four ANOVAs were performed with either NART or WTAR at wave 1 or 2 as dependent variables and testers as fixed factors. There were significant tester effects on NART and WTAR scores at wave 1 and on NART at wave 2. The effect sizes were relatively small, η^2 ranging from 0.009 to 0.059, with only one (NART at wave 1) in the medium effect size range (Cohen, 1988). Post-hoc comparisons for the largest effect revealed that participants tested by one particular tester had on average higher NART scores than participants tested by other testers. To control for the possibility that there were systematic differences in the ability of individuals tested by different testers, we controlled for cognitive ability at age 70 (MHT score). Differences in the MHT scores did not explain tester effects.

We also calculated intra- and inter-rater reliability on a subsample of raters, for which a two-way random model ICC could be used, i.e. those that were involved in both waves of testing. We calculated intra-rater reliability for the two raters who administered NART and WTAR to the same participants at both testing waves. These ICCs were high for both NART (.93, .95) and for WTAR (.88, .96). For comparison, we also obtained inter-rater reliability estimates for the same two raters in the cases where one of them tested a given participant at

SUPPLEMENT

wave 1 and the other one at wave 2. These were also very high for both NART scores (.92, .96) and WTAR scores (.89, .94). Therefore inter-rater agreement was high for both of these reading test scores and comparable with intra-rater agreement.

Table S1. Mean National Adult Reading Test (NART) and Wechsler Test of Adult Reading (WTAR) scores by different testers

| Tester | Test | Wave 1 | | | Wave 2 | | | IRR |
|----------------|------|--|-------|-----|---|-------|------------------|-----|
| | | Mean (SD) | Range | N | Mean (SD) | Range | N | |
| 2 | NART | 34.98 (6.95) | 15-48 | 95 | - | - | - | |
| | WTAR | 42.28 (5.63) | 25-50 | | - | - | - | |
| 3 | NART | 33.77 (8.58) | 12-48 | 171 | - | - | - | |
| | WTAR | 39.32 (7.82) | 19-50 | | - | - | - | |
| 5 | NART | 39.28 (6.73) | 13-50 | 123 | - | - | - | |
| | WTAR | 44.28 (5.86) | 16-50 | | - | - | - | |
| 1 | NART | 35.44 (7.15) | 20-49 | 137 | 36.41 (7.24) | 14-49 | .93 ^a | |
| | WTAR | 41.74 (6.18) | 22-50 | | 42.43 (6.28) | 21-50 | .96 ^a | |
| 4 | NART | 33.79 (7.85) | 14-49 | 276 | 33.62 (7.95) | 10-49 | .95 ^b | |
| | WTAR | 41.55 (6.51) | 22-50 | | 40.48 (6.83) | 18-50 | .88 ^b | |
| 6 | NART | - | - | | 34.48 (8.44) | 10-49 | - | |
| | WTAR | - | - | | 41.79 (7.12) | 18-50 | - | |
| 7 | NART | - | - | | 36.00 (6.54) | 21-48 | - | |
| | WTAR | - | - | | 41.23 (5.69) | 28-50 | - | |
| 8 | NART | - | - | | 34.63 (7.90) | 9-50 | - | |
| | WTAR | - | - | | 41.40 (6.47) | 21-50 | - | |
| Tester effects | | | | | | | | |
| | NART | F(4,797)=12.59, p<.001, eta ² =.059 | | | F(4,797)=2.84, p=.023, eta ² =.014 | | | |
| | WTAR | F(4,797)=10.58, p<.001, eta ² =.050 | | | F(4,797)=1.83, p=.121, eta ² =.009 | | | |

Note. IRR=intra-rater reliability

^a Calculation based on n=8

^b Calculation based on n=89