Journal of Clinical and Experimental Neuropsychology

Functional equivalence of the National Adult Reading Test (NART) and Schonell reading tests and NART norms in the Dynamic Analyses to Optimise Ageing (DYNOPTA) project

Kim M. Kiely a, Mary A. Luszcz b, Olivier Piguet c d, Helen Christensen a, Hayley Bennett c & Kaarin J. Anstey a

a Centre for Mental Health Research, The Australian National University, Australia
b Flinders University, Adelaide, South Australia, Australia
c Prince of Wales Medical Research Institute, Sydney, New South Wales, Australia
d School of Medical Sciences, the University of New South Wales, Sydney, New South Wales, Australia

Published online: 04 Dec 2010.

To cite this article: Kim M. Kiely, Mary A. Luszcz, Olivier Piguet, Helen Christensen, Hayley Bennett & Kaarin J. Anstey (2011) Functional equivalence of the National Adult Reading Test (NART) and Schonell reading tests and NART norms in the Dynamic Analyses to Optimise Ageing (DYNOPTA) project, Journal of Clinical and Experimental Neuropsychology, 33:4, 410-421, DOI: 10.1080/13803395.2010.527321

To link to this article: http://dx.doi.org/10.1080/13803395.2010.527321

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Functional equivalence of the National Adult Reading Test (NART) and Schonell reading tests and NART norms in the Dynamic Analyses to Optimise Ageing (DYNOPTA) project

Kim M. Kiely1, Mary A. Luszcz2, Olivier Piguet3,4, Helen Christensen1, Hayley Bennett3, and Kaarin J. Anstey1

1Centre for Mental Health Research, The Australian National University, Australia
2Flinders University, Adelaide, South Australia, Australia
3Prince of Wales Medical Research Institute, Sydney, New South Wales, Australia
4School of Medical Sciences, the University of New South Wales, Sydney, New South Wales, Australia

This study investigates the functional equivalence of two measures of irregular word pronunciation—National Adult Reading Test (NART) and Schonell—which are popular instruments used to assess verbal neurocognitive functioning and to estimate premorbid IQ. We report norms for the NART in a pooled sample from 3 Australian population-based studies of adults aged 65–103 years. Norms were stratified by sex and age left school in 5-year age groups. The NART and the Schonell had a strong linear relation, allowing for the imputation of NART scores based on Schonell performance within 1 study. Neither measure was sensitive to the effects of sex after adjusting for the effects of age and education. Early school leavers performed worse on both measures. Data pooling enables greater precision and improved generalizability of NART norms than do methods that use single older adult samples.

Keywords: Verbal abilities; Harmonization; Premorbid IQ; Crystallized intelligence; National Adult Reading Test; Norms; Dynamic Analyses to Optimise Ageing; Australian Longitudinal Study of Ageing; Canberra Longitudinal Study of Ageing; Sydney Older Persons Study.

Investments in large-scale studies of neuropsychological function have underwritten the current movement towards a greater synthesis and optimal use of existing data sources within ageing research disciplines. One strategy to achieve synthesis is data pooling and variable harmonization. Pooling independently designed studies enables analyses of large datasets, increases statistical power to investigate low-prevalence disorders, aids comparative cross-population research, and provides for simultaneous replication of empirical findings (Anstey et al., 2010; Hofer & Piccinin, 2009; van Buuren, Eyres, Tennant, & Hopman-Rock, 2005). Data...
pooling is particularly useful for the generation of norms, which otherwise are often based on small sample studies with restricted population coverage (e.g., Tombaugh & Hubiey, 1997; Uttl, 2002). Although the notion of multistudy analysis via data pooling or meta-analysis is not new, often pooled studies will not share operationally identical measures. In these instances, data pooling is contingent upon the existence of measures that index the same (though not identical) theoretical construct and so are functionally equivalent. A growing literature explicitly addresses the use of harmonization methods for combining functionally equivalent measures to facilitate multistudy analyses (McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009; Pommerich & Dorans, 2004; van Buuren et al., 2005). Research on the harmonization of cognitive measures is a relatively recent endeavor despite the more developed practice of data pooling of biological variables, such as serum cholesterol, in other epidemiological and medical disciplines (e.g., Dyer, 1986).

Both the Schonell Graded Word Reading Test (“Schonell”; Schonell, 1942) and the National Adult Reading Test (NART; Nelson, 1982) are 50-item graded pronunciation tasks for irregularly pronounced words. They each assess verbal abilities and have been used as a proxy for premorbid verbal IQ (VIQ) and full-scale IQ (FSIQ) in adults (Crawford, Deary, Starr, & Whalley, 2001; Crawford, Parker, Stewart, Besson, & De Lacey, 1989; Nelson & McKenna, 1975) based on the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). The Schonell, however, was originally developed to assess reading ability in children and is therefore relatively insensitive to reading ability in well-educated adults (Franzen, 2000). In order to provide more accurate indexation of premorbid IQ, the NART was developed and standardized in an adult population in Britain (Nelson, 1982; Nelson & McKenna, 1975) and has since been revised (NART–2) for the Wechsler Adult Intelligence Scale–Revised (WAIS–R) and for North American populations (American National Adult Reading Test, AMNART; North American Adult Reading Test, NAART). Short forms have also been developed (Strauss, Sherman, & Spreen, 2006).

The NART has robust reliability, sound psychometric properties, and brief administration time, and it is resistant against cognitive decline except in severe cases or in the presence of focal reading disorders or semantic dementia. Reflecting their strong indexation of putative crystallized abilities (Horn & Cattell, 1967), longitudinal analyses of word pronunciation tasks have revealed their relative stability in late life, typically showing only slight age-related declines in performance with no individual differences in intraindividual variability (Anstey, Hofer, & Luszcz, 2003). For these reasons, the NART is commonly used as a proxy measure of premorbid IQ in large population-based studies used in epidemiological and cognitive neurobiological research (e.g., Anstey et al., 2003; Christensen et al., 1999; Luszcz, Bryan, & Kent, 1997).

The NART and the Schonell are functionally equivalent measures that share similar task parameters and differ predominantly in item difficulty. Research investigating their shared psychometric properties has primarily been in relation to predicting WAIS IQ. Crawford et al. (1989), for example, investigated whether combining the NART and Schonell tasks would provide increased prediction accuracy of IQ for subjects with limited verbal abilities (i.e., NART items correct <10). They reported, however, that doing so systematically underestimated WAIS IQ and resulted in a nonlinear relationship between predicted and observed WAIS IQ. There has been little research investigating the direct association and compatibility between these two measures in nonclinical populations.

The available norms for the NART itself are limited. NART norms have not been derived from random samples of the general population, but rather have been generated through recruitment via interest groups, advertisements, and convenience samples. Sample size restrictions have also necessitated the reporting of normative data for the NART in overlapping 10-year age groups, with typically small cell counts for older age groups (e.g., Spreeen & Strauss, 1998; Uttl, 2002). For example, norms reported for adults between the ages of 50 and 100 years, stratified by 5-year age groups, sex, and a binary measure of education, would result in 40 subgroups. The likelihood of low cell counts in these subgroups brings into question the generalizability of reported norms, particularly for older age groups, and undermines the validity of regression-based norms that adjust for sociodemographic variables (Fastenau, 1998). Moreover, NART norms specific to the Australian population do not exist. Currently, Australian clinicians must use UK NART norms, which may be biased by cultural differences in curriculum, vernacular, and pronunciation (Mathias, Bowden, & Barrett-Woodbridge, 2007).

**Objectives**

The objectives of this study were twofold. The first aim was to assess the functional equivalence of NART and Schonell tasks and the capacity to produce comparable predicted NART scores based on Schonell performance within a single study where baseline NART scores were not obtained. The second aim was to perform multistudy comparisons across three large population-based studies that have obtained NART data. The baseline data from these studies were used to produce normative NART data based on a sample of over 2,636 older Australians, producing greater subgroup precision than previously reported. The relations between the NART and contextual variables were also explored.

**METHOD**

**Study design**

The data were drawn from the Dynamic Analyses to Optimise Ageing (DYNOPTA) project (Anstey et al,
dataset, which has harmonized and pooled data from nine Australian longitudinal studies of ageing. Three studies within DYNOPTA administered either the NART or the Schonell. Both the Australian Longitudinal Study of Ageing (ALSA; Luszcz et al., 2007) and the Canberra Longitudinal Study (CLS; Christensen et al., 2004) have repeated measures of the NART on up to four occasions. The Sydney Older Person Study (SOPS) included the Schonell at both baseline and first follow-up, whereas the NART was only included at the first follow-up assessment. Baseline NART scores are therefore study censored for SOPS participants. Study censoring refers to data that were not obtained by a particular study. This is a common problem encountered in data pooling. Because the NART was not administered at baseline, the mechanism for study-censored data is known and can be classified entirely as Missing at Random (MAR; Little & Rubin, 2002; Salthouse, 2004). The first study aim relates to predicting baseline NART scores for this study-censored population by imputation procedures that are appropriate for data that are assumed MAR.

Contributing study sample profiles

The study-censored SOPS sample consisted of a random sample of community-dwelling residents aged 75 years and older from eight local government areas from the inner west of Sydney. The two-stage sampling design included 327 World War II veterans and widows and 320 nonveterans. Because of the selection process, 17 participants were sampled in both sampling frames (Piguet et al., 2003). Clinical interview and cognitive assessments were conducted by a trained physician experienced in geriatric medicine. The investigation of the functional equivalence between the NART and Schonell is based on data from the first follow-up wave when each measure was obtained concurrently in SOPS.

Within SOPS the baseline measurement wave occurred between August of 1991 and September of 1993 and included 630 participants (females = 50.5%) with a mean age of 80.5 years (SD = 4.2). The average time interval between baseline and the first follow-up wave was 2.9 years (SD = 0.3). A total of 128 participants died before the start of the first follow-up wave; 448 respondents (female = 54.7%) participated in the first follow-up wave, resulting in an attrition rate of 12.9%. The average school leaving age was 14.6 years (SD = 5.9, range: 64–103), and average score on the Mini-Mental State Examination (MMSE) was 27.1 (SD = 2.9). Baseline waves from the contributing studies were contemporaneous, and NART data were obtained between the years 1990 and 1993.

Pooled sample profile

Pooled baseline data from the three studies were used to generate NART norms, which are based on observed NART scores for ALSA and CLS and predicted NART scores for SOPS. The pooled sample (see Table 1) comprised 2,636 participants (female = 51.1%). Of these, 48.2% left secondary school at the age of 14 years or younger, 73% were born in Australia, and 95% reported English as their first language. The average age was 77.8 years (SD = 5.9, range: 64–103), and average score on the Mini-Mental State Examination (MMSE) was 27.1 (SD = 2.9). Baseline waves from the contributing studies were contemporaneous, and NART data were obtained between the years 1990 and 1993.

Measures

In SOPS, the Schonell was administered at both baseline and the first follow-up wave, with the NART being administered in the first follow-up wave only. The ALSA and the CLS each administered the NART on four measurement occasions; however, only baseline NART scores are used for the generation of norms. The NART and Schonell were each scored as the number of words correctly pronounced and ranged between 0 and 50.

Time-invariant covariates were obtained at baseline and included “sex” (0 = male, 1 = female), “language first spoken” (0 = English, 1 = other), and “age left school.” A continuous measure of age left school was used to test whether education mediated the relation between the NART and Schonell in SOPS. To ensure commonality across studies when reporting norms, a harmonized variable for “age left school” was collapsed to a binary coding (0 = 14 years or younger; 1 = 15 years or older). Time-varying covariates included “age at time of observation” and MMSE scores (Folstein, Folstein, & McHugh, 1975), which were used as an index of cognitive status. Covariates were chosen as candidates for inclusion in the imputation model because of their documented relations with the NART and Schonell.

Analysis

Imputation of NART scores in SOPS

Baseline NART scores were imputed by conditional ordinary least squares mean imputation (Little & Rubin 2002; McKnight, McKnight, Sidani, & Figueredo, 2007). Bivariate Pearson correlations were used to determine association between the Schonell, NART, and candidate covariates. Covariates were mean centered to minimize multicollinearity. Backwards hierarchical regression was used to identify predictor covariates for inclusion in the imputation model, with both linear and nonlinear effects. Backwards regression is appropriate because the aim of this exploratory analysis is to identify the optimal model for predicting NART scores, rather than hypothesis testing. The NART was not imputed for cases.
<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>&lt;70 (n = 96)</th>
<th>70–74 (n = 740)</th>
<th>75–79 (n = 839)</th>
<th>80–84 (n = 577)</th>
<th>85–89 (n = 286)</th>
<th>90–94 (n = 80)</th>
<th>≥ 95 (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>10 (0.00)</td>
<td>387 (0.15)</td>
<td>452 (0.17)</td>
<td>299 (0.11)</td>
<td>161 (0.06)</td>
<td>34 (0.01)</td>
<td>4 (0.00)</td>
</tr>
<tr>
<td>Females</td>
<td>86 (0.03)</td>
<td>353 (0.13)</td>
<td>387 (0.15)</td>
<td>278 (0.11)</td>
<td>125 (0.05)</td>
<td>46 (0.02)</td>
<td>14 (0.01)</td>
</tr>
<tr>
<td>City</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adelaide</td>
<td>96 (0.04)</td>
<td>369 (0.14)</td>
<td>335 (0.13)</td>
<td>252 (0.10)</td>
<td>170 (0.06)</td>
<td>35 (0.01)</td>
<td>7 (0.00)</td>
</tr>
<tr>
<td>Canberra</td>
<td>0 (0.00)</td>
<td>371 (0.14)</td>
<td>267 (0.10)</td>
<td>171 (0.06)</td>
<td>59 (0.02)</td>
<td>32 (0.01)</td>
<td>7 (0.00)</td>
</tr>
<tr>
<td>Sydney</td>
<td>0 (0.00)</td>
<td>0 (0.00)</td>
<td>237 (0.09)</td>
<td>154 (0.06)</td>
<td>57 (0.02)</td>
<td>13 (0.00)</td>
<td>4 (0.00)</td>
</tr>
<tr>
<td>Career occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managers and professionals</td>
<td>7 (0.00)</td>
<td>203 (0.01)</td>
<td>204 (0.10)</td>
<td>140 (0.07)</td>
<td>77 (0.04)</td>
<td>20 (0.01)</td>
<td>2 (0.00)</td>
</tr>
<tr>
<td>Clerical and associate professional</td>
<td>18 (0.01)</td>
<td>155 (0.07)</td>
<td>134 (0.06)</td>
<td>86 (0.04)</td>
<td>37 (0.02)</td>
<td>14 (0.01)</td>
<td>4 (0.00)</td>
</tr>
<tr>
<td>Tradespersons</td>
<td>14 (0.01)</td>
<td>107 (0.05)</td>
<td>103 (0.05)</td>
<td>92 (0.04)</td>
<td>46 (0.02)</td>
<td>5 (0.00)</td>
<td>1 (0.00)</td>
</tr>
<tr>
<td>Sales, service, production, transport, laborers</td>
<td>14 (0.01)</td>
<td>141 (0.07)</td>
<td>245 (0.12)</td>
<td>127 (0.06)</td>
<td>57 (0.03)</td>
<td>26 (0.01)</td>
<td>7 (0.00)</td>
</tr>
<tr>
<td>Age left school (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤14</td>
<td>46 (0.02)</td>
<td>343 (0.13)</td>
<td>387 (0.15)</td>
<td>300 (0.11)</td>
<td>151 (0.06)</td>
<td>35 (0.01)</td>
<td>8 (0.00)</td>
</tr>
<tr>
<td>≥15</td>
<td>50 (0.02)</td>
<td>395 (0.15)</td>
<td>449 (0.17)</td>
<td>274 (0.10)</td>
<td>135 (0.05)</td>
<td>43 (0.02)</td>
<td>9 (0.00)</td>
</tr>
<tr>
<td>Qualifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>64 (0.03)</td>
<td>404 (0.16)</td>
<td>533 (0.21)</td>
<td>349 (0.14)</td>
<td>164 (0.06)</td>
<td>46 (0.02)</td>
<td>11 (0.00)</td>
</tr>
<tr>
<td>Postsecondary, Nontertiary</td>
<td>25 (0.01)</td>
<td>255 (0.10)</td>
<td>218 (0.09)</td>
<td>163 (0.06)</td>
<td>82 (0.03)</td>
<td>22 (0.01)</td>
<td>4 (0.00)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>1 (0.00)</td>
<td>65 (0.03)</td>
<td>63 (0.02)</td>
<td>34 (0.01)</td>
<td>20 (0.01)</td>
<td>7 (0.00)</td>
<td>1 (0.00)</td>
</tr>
<tr>
<td>Country of birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born in Australia</td>
<td>62 (0.02)</td>
<td>522 (0.20)</td>
<td>650 (0.25)</td>
<td>426 (0.16)</td>
<td>197 (0.07)</td>
<td>54 (0.02)</td>
<td>14 (0.01)</td>
</tr>
<tr>
<td>Not born in Australia</td>
<td>34 (0.01)</td>
<td>218 (0.08)</td>
<td>189 (0.07)</td>
<td>150 (0.06)</td>
<td>89 (0.3)</td>
<td>26 (0.01)</td>
<td>4 (0.00)</td>
</tr>
<tr>
<td>Language first spoken</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>93 (0.04)</td>
<td>685 (0.26)</td>
<td>801 (0.31)</td>
<td>550 (0.21)</td>
<td>280 (0.11)</td>
<td>76 (0.03)</td>
<td>18 (0.01)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (0.00)</td>
<td>52 (0.02)</td>
<td>24 (0.01)</td>
<td>6 (0.00)</td>
<td>3 (0.00)</td>
<td>0 (0.00)</td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td></td>
<td>28.5 (2.17)</td>
<td>28.1 (1.99)</td>
<td>27.2 (2.75)</td>
<td>26.6 (3.08)</td>
<td>25.8 (3.09)</td>
<td>25.1 (3.43)</td>
<td>22.1 (5.24)</td>
</tr>
</tbody>
</table>

Note. N = 2,636. MMSE: Mini-Mental State Examination.
where concurrent Schonell data were unavailable; this preserves missing data patterns that are conditional on factors other than survey design, such as respondent characteristics. Conditional ordinary least squares mean imputation is attractive because it is simple, produces a single estimate, is based on existing relations in the data, and provides more accurate estimates than unconditional mean imputation or follow-up NART scores carried backward. Stochastic regression or multiple imputation, however, remain preferable as they provide more robust imputation procedures with closer approximations of the conditional distributions and, in the case of multiple imputation, account for the uncertainty of predicted values (Graham, 2009; Little & Rubin, 2002; Schafer & Graham, 2002). For this reason, results of the conditional mean imputation procedure will be compared with multiple imputation of five datasets.

Criteria for assessing imputed values

Criteria for the most reliable estimates were the fit of distribution properties, the shared variance between predicted and observed NART scores, and the time-lagged associations between the NART and Schonell within the SOPS sample.

Generation of norms

NART norms were reported for subgroups stratified by 5-year birth cohorts, sex, and a binary measure of age left school using harmonized baseline data from ALSA, CLS, and SOPS. Analysis of covariance (ANCOVA) was used to test for study effects and to investigate the relation between the NART and sociodemographic variables.

Evaluation of age differences and age changes

Mixed models (Singer & Willett, 2003) were used to estimate annualized within-person change in NART scores to provide a comparison with age differences in NART performance based on cross-sectional data. Unadjusted and adjusted (controlling for study, sex, age left school, and total MMSE) estimates were computed using STATA 10 (2007).

RESULTS

Baseline NART prediction for SOPS

At follow-up for SOPS, NART scores ($M = 25.96, SD = 9.98$) were lower than the Schonell scores ($M = 35.47, SD = 10.7$). The Schonell scores were not normally distributed and were negatively skewed at both waves (baseline: $-1.08$; follow-up: $-1.01$). To correct for the negative skew, a square root transformation was applied. After the transformation, Schonell scores had a skew of 0.13 ($SE = 0.11$) at baseline and 0.36 ($SE = 0.13$) at follow-up.

No sex differences in Schonell scores were observed at baseline (mean difference $= 1.4; p = .14$). At follow-up, however, males performed better than females on both the NART (mean difference $= 3.3, p < .001$) and Schonell (mean difference $= 2.8, p = .02$). This difference in scores was thought to reflect the greater propensity of lower functioning males to drop out of the study due to morbidity or mortality-related factors. We therefore investigated sociodemographic differentials for mortality and attrition. Chi square test revealed no sex differences in attrition not associated with mortality ($\chi^2 = 0.01, p = .94$); however, males were more likely to die between baseline and first follow-up ($\chi^2 = 13.9, p < .001$). We also regressed a binary indicator of mortality prior to the first follow-up wave on age, sex, MMSE, and age left school. Participants who died between baseline and first follow-up were more likely to be older (odds ratio, $OR = 1.1, SE = .03, p < .001$), male ($OR = 2.9, SE = .80, p < .001$) and score below 23 on the MMSE ($OR = 1.8, SE = .57, p = .038$). Age left school was not associated with mortality between baseline and first follow-up ($OR = 1.2, SE = .30, p = .41$).

Relation between the NART and Schonell

The Schonell scores were strongly associated with NART scores ($r = .913, p < .001$; Figure 1). The scatter plot in Figure 1 also revealed ceiling effects for the Schonell, particularly for individuals with more years of schooling. Table 2 shows the bivariate correlations between the NART and candidate covariates for the predictor model. Due to the strong association between the Schonell and the NART, the effects of age, sex, and MMSE on NART scores were not significant and were dropped from the final imputation model. It should not be inferred that these covariates are not predictive of NART performance; rather, this is an indication of the functional equivalence between the NART and Schonell. The exclusion of age, sex, and MMSE from the imputation model is desirable as future analyses that model associations between these covariates and verbal abilities indexed by the NART will not produce inflated covariances. At follow-up, the main effects of Schonell scores (linear and quadratic components) and age left school accounted for 87% of variability in NART scores.

SOPS baseline NART prediction from Schonell scores

Only follow-up NART scores, time-varying Schonell scores, and a continuous measure of age left school were retained in the final imputation model used to predict baseline NART scores for SOPS participants. Eight imputation models were compared, and all explained between 83.9% and 86.8% of variability in NART scores. Models that included age left school imputed baseline NART scores for 465 participants and imputed follow-up NART scores for 3 participants. Models that excluded age left school from the model imputed baseline NART scores for 476 participants and imputed
follow-up NART scores for 8 participants. Models that did not correct for the skew in Schonell scores systematically underestimated NART scores for early school leavers and also appeared to be a poor fit of the distribution properties for the NART. Models that excluded the quadratic effect overestimated ability levels of individuals with low NART scores at follow-up. The final model was selected as it showed the smallest standard error ($SE = 3.65$) and showed the closest match to the distribution of observed NART scores at follow-up (Figure 2, Table 3).

The predicted NART scores were not rounded to the nearest whole number and were imputed by the unstandardized regression equation:

$$NART_{\text{pred}} = (20.651) + (-6.225) \times (\text{Schonell}_{\text{trans}})$$

$$+ (-.225) \times (\text{Schonell}_{\text{trans}}^2) + (.399) \times (\text{als}),$$

where $\text{Schonell}_{\text{trans}}$ is the mean centred transformed Schonell score, $\text{Schonell}_{\text{trans}}^2$ is the quadratic effect for the mean centered transformed Schonell score, and $\text{als}$ is age left school. There was an average decline of 2.39

**TABLE 2**

Pairwise bivariate correlations between the NART, Schonell, and covariates at follow-up for SOPS

<table>
<thead>
<tr>
<th></th>
<th>NART</th>
<th>Schonell</th>
<th>MMSE</th>
<th>ALScont</th>
</tr>
</thead>
<tbody>
<tr>
<td>NART</td>
<td>1</td>
<td>.91**</td>
<td>.52**</td>
<td>.31**</td>
</tr>
<tr>
<td>Schonell</td>
<td>.91**</td>
<td>1</td>
<td>-.93**</td>
<td>-.52**</td>
</tr>
<tr>
<td>MMSE</td>
<td>.52**</td>
<td>-.93**</td>
<td>1</td>
<td>.03</td>
</tr>
<tr>
<td>ALScont</td>
<td>.31**</td>
<td>-.52**</td>
<td>.03</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>-.22**</td>
<td>-.97**</td>
<td>-.22**</td>
<td>-.09</td>
</tr>
</tbody>
</table>

Note. $N = 345$. MMSE: Mini-Mental State Examination. NART: National Adult Reading Test. SOPS: Sydney Older Persons Study. ALScont: age left school, continuous measure; Schonellraw: raw Schonell scores; Schonelltrans: square-root transformed Schonell scores.

*$p < .01$, two-tailed.

Figure 1. Relation between NART (National Adult Reading Test) and Schonell scores. Scatter plot of raw NART and Schonell scores by age left school for SOPS (Sydney Older Persons Study) at Wave 2 ($r = .91$, $p < .001$). The relationship is best modeled with both a linear and a quadratic component ($r = .93$, $p < .001$).
points ($SD = 4.81$) in NART performance between baseline and follow-up ($M_{\text{base}} = 2.9, SD = 0.3$), compared to an average decline of 2.61 points ($SD = 4.77$) on the Schonell. The correlation between the NART and Schonell difference scores was $r = .61 (p < .001)$, indicating a moderate to strong association in change over two waves between these two measures.

The variances and correlations between predicted and observed NART and Schonell scores across both waves are presented in Table 4. Correlations ranged between .85 and .98. No significant differences were found between predicted baseline NART scores estimated by imputation ($M = 28.22, SE = 0.44$) or multiple imputation ($M = 28.11, SE = 0.53$), $t(10) = 0.363, p = .725; r = .93, p < .001$. It is evident that the final mean imputation model underestimated variability, as predicted NART scores had a standard deviation of 9.38 at follow-up, whereas the actual NART standard deviation was 10.4. Further, the mean standard error was larger for estimates derived from multiple imputation. This was not unexpected and is a common problem of conditional mean imputation as predicted values do not deviate from the regression line (Graham, 2009). Baseline NART norms for the SOPS sample are based on the values imputed by conditional mean imputation.

### NART norms

**Normative data**

Only the CLS provided item-level data that could be used to investigate the psychometric properties of the NART. Reliability analysis for this sample revealed that Cronbach’s alpha was comparable across all age groups and ranged from .85 to .93 (Table 5). This is generally consistent with previous findings. For example Uttl (2002) reported alphas of .92, .94, and .93 for broad age cohorts (young, 18–39 years; middle-aged, 40–59 years; and older adults, 60–91 years, respectively). The relatively lower alpha of .85 reported for this sample was for the oldest age group (95+), which had a small sample size ($n = 18$) and comprised an age cohort not captured in the Uttl (2002) sample. Luszcz et al. (1997) reported test–retest reliability of .83 across all ages ranging from 65 to 103 years. Test–retest reliability is considered a more

---

**Figure 2.** Relation between predicted and observed NART (National Adult Reading Test) scores. Scatter plot of predicted (y-axis) and observed NART (x-axis) scores for SOPS (Sydney Older Persons Study) Wave 2 ($r = .93, p < .001$).
conservative reliability estimate relative to measures of internal consistency such as Cronbach’s alpha.

Table 5 shows the normative data at baseline for the NART in five-year age groups by sex and age left school. ANCOVA revealed that NART performance was related to age left school and qualifications attained. Individuals who left school before the age of 15 years (M = 25.46, SD = 9.22, n = 1,270) performed worse than participants who left school at ages 15 years or older (M = 31.74, SD = 9.69, n = 1,355; p < .001). A negative association was also present between age and NART scores: Every 10 years lived resulted in a predicted difference of –1 point in NART score (p < .001). No sex effects were found (p = .47) after adjusting for age and education. Care must be taken when interpreting age differences in NART performance as this finding is based on cross-sectional analysis, which cannot distinguish between age and cohort effects, nor permit inference concerning within-person change. To examine this further, subsequent longitudinal analysis of the pooled sample tested...
TABLE 5

Normative baseline NART scores from the pooled DYNOPTA dataset by 5-year age group, sex, and age left school

<table>
<thead>
<tr>
<th>Age group</th>
<th>&lt;70 (n = 96)</th>
<th>70–74 (n = 740)</th>
<th>75–79 (n = 839)</th>
<th>80–84 (n = 577)</th>
<th>85–89 (n = 286)</th>
<th>90–94 (n = 80)</th>
<th>≥95 (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>.92</td>
<td>.93</td>
<td>.92</td>
<td>.92</td>
<td>.91</td>
<td>.91</td>
<td>.85</td>
</tr>
<tr>
<td>Whole sample</td>
<td>n = 96</td>
<td>n = 740</td>
<td>n = 839</td>
<td>n = 577</td>
<td>n = 286</td>
<td>n = 80</td>
<td>n = 18</td>
</tr>
<tr>
<td>M (SD)</td>
<td>27.3 (8.6)</td>
<td>29.7 (9.92)</td>
<td>28.9 (10.19)</td>
<td>28.2 (9.82)</td>
<td>27.1 (9.73)</td>
<td>27.3 (10.84)</td>
<td>25.5 (9.63)</td>
</tr>
<tr>
<td>Range</td>
<td>4.45</td>
<td>0–50</td>
<td>0–49</td>
<td>0–50</td>
<td>0–50</td>
<td>0–50</td>
<td>9–40</td>
</tr>
<tr>
<td>Males ALS ≤ 14 years</td>
<td>n = 5</td>
<td>180</td>
<td>217</td>
<td>176</td>
<td>89</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>M (SD)</td>
<td>20.6 (10.31)</td>
<td>27.0 (9.52)</td>
<td>25.6 (10.1)</td>
<td>26.1 (8.9)</td>
<td>24.2 (9.47)</td>
<td>21.2 (11.34)</td>
<td>24.0 (11.31)</td>
</tr>
<tr>
<td>Range</td>
<td>9–32</td>
<td>0–46</td>
<td>0–44</td>
<td>2–50</td>
<td>4–50</td>
<td>0–36</td>
<td>16–32</td>
</tr>
<tr>
<td>ALS ≥ 15 years</td>
<td>n = 5</td>
<td>207</td>
<td>234</td>
<td>122</td>
<td>72</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>M (SD)</td>
<td>25.4 (7.23)</td>
<td>33.4 (9.76)</td>
<td>32.3 (9.74)</td>
<td>32.0 (10.19)</td>
<td>29.9 (9.52)</td>
<td>29.8 (8.56)</td>
<td>33.5 (6.04)</td>
</tr>
<tr>
<td>Range</td>
<td>18–35</td>
<td>0–50</td>
<td>0–48</td>
<td>2–50</td>
<td>10–47</td>
<td>13–50</td>
<td>29–38</td>
</tr>
<tr>
<td>Total</td>
<td>M (SD)</td>
<td>23 (8.77)</td>
<td>30.4 (10.16)</td>
<td>29.1 (10.46)</td>
<td>28.5 (9.87)</td>
<td>26.8 (9.88)</td>
<td>25.2 (10.77)</td>
</tr>
<tr>
<td>Females ALS ≤ 14 years</td>
<td>n = 41</td>
<td>163</td>
<td>170</td>
<td>124</td>
<td>62</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>M (SD)</td>
<td>23.6 (7.71)</td>
<td>25.7 (8.53)</td>
<td>25.3 (9)</td>
<td>24.8 (8.57)</td>
<td>24.9 (9.59)</td>
<td>24.4 (7.41)</td>
<td>25.0 (9.61)</td>
</tr>
<tr>
<td>Range</td>
<td>4–38</td>
<td>1–45</td>
<td>0–44</td>
<td>1–50</td>
<td>0–42</td>
<td>9–40</td>
<td>14–40</td>
</tr>
<tr>
<td>ALS ≥ 15 years</td>
<td>n = 45</td>
<td>188</td>
<td>215</td>
<td>152</td>
<td>63</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>M (SD)</td>
<td>31.6 (7.36)</td>
<td>32.0 (9.4)</td>
<td>31.5 (9.78)</td>
<td>30.7 (9.9)</td>
<td>30.1 (8.86)</td>
<td>31.4 (11.79)</td>
<td>26.1 (10.04)</td>
</tr>
<tr>
<td>Range</td>
<td>18–45</td>
<td>5–49</td>
<td>6–49</td>
<td>6–50</td>
<td>11–49</td>
<td>1–47</td>
<td>9–39</td>
</tr>
<tr>
<td>Total</td>
<td>M (SD)</td>
<td>27.8 (8.49)</td>
<td>29 (9.61)</td>
<td>28.8 (9.89)</td>
<td>27.9 (9.78)</td>
<td>27.5 (9.56)</td>
<td>28.9 (10.75)</td>
</tr>
</tbody>
</table>

Note. Cronbach’s alpha based on the Canberra sample only. Test-retest reliability has previously been reported in the Adelaide sample across all age groups to be .83 (Luszcz et al., 1997). NART: National Adult Reading Test. DYNOPTA: Dynamic Analyses to Optimise Ageing. ALS: age left school.
for age trajectories of intraindividual change in NART score. Participants were assessed on the NART on up to four occasions (M = 1.9), and the average time interval between waves was 2.2 years (SD = 0.7; range 0.09–5.9). Unadjusted linear mixed models revealed an average decline in NART performance of 1.4 words correctly pronounced over 10 years (SE = 0.17, p < .001). However, after adjusting for study, sex, age left school (binary), and total MMSE score, the average rate of decline in NART attenuated to 0.4 words correctly pronounced over 10 years (SE = 0.17, p = .02).

Study differences

Differences in baseline NART performance across studies were observed: The Canberra sample (M = 30.5, SD = 11.7) scored higher than both Sydney (M = 28.2, SD = 9.58, p < .001) and Adelaide (M = 27.52, SD = 8.48, p < .001) samples. After adjusting for age, sex, age left school, qualifications attained, language first spoken, and MMSE, study differences were no longer present between the estimated marginal means for the Sydney and Canberra samples (p = .75), although lower performance for Adelaide participants remained (p < .001).

DISCUSSION

This study investigated the functional equivalence of two irregular word pronunciation tests and found that the Schonell and NART have a strong linear relation, with a slight curvilinear component. These findings indicate that NART scores can be estimated based on the Schonell and age left school. Given the strength of the relation between the tests, little benefit is gained from administering each measure concurrently aside from the opportunity to investigate their functional equivalence.

Imputation of baseline NART scores for SOPS participants ensures that NART norms reported for this population are not biased by participative effects, attrition, or mortality. Although verbal abilities indexed by the NART are not expected to decline with age, NART performance may decline in relation to disease or mortality processes (White & Cunningham, 1988). Thus, baseline estimates also lay the groundwork for future investigations into individual differences in change of NART scores. Salthouse (2009) has recently argued that the true extent of age-related declines in cognition will be largely obscured in longitudinal studies that do not account for learning effects, to which verbal tasks are highly susceptible. The use of alternate measures is one approach to minimizing these participative effects. This paper therefore joins a growing realization that identical measures need not be applied across all measurement occasions of longitudinal studies. Longitudinal analyses can be applied to changing scales of functionally equivalent measures (McArdle et al., 2009) and benefit from planned missingness designs (Graham, Taylor, & Cumsille, 2001). Although in this instance the opportunity to investigate the functional equivalence of two measures of verbal abilities was largely serendipitous, researchers involved in the design of future longitudinal studies may wish to consider incorporating alternate forms of functionally equivalent measures across waves.

The second aim of this study was to combine three large population-based studies to report normative NART data with greater precision and representation of a broader national population. This is a key strength of the study as historically normative NART data have been based on relatively small samples of recruited volunteers or from community groups via advertisements and are not drawn from a random sample of the general population.

Our investigations confirmed that educational attainment is related to NART scores: Early school leavers systematically performed worse than participants who completed secondary schooling. Likewise, those failing to obtain a postsecondary schooling qualification were more likely to incorrectly pronounce irregular words than individuals who did obtain further postsecondary and tertiary qualifications. Sex did not account for any variability in NART scores after adjusting for sociodemographics, which is consistent with previous findings. Contrary to reported norms for the North American Adult Reading test (NAART), which show improved NART performance in older adults (Strauss et al., 2006; Uttl, 2002), older adults were more likely to perform worse than their younger counterparts in this study. This difference, however, is most likely due to the greater age range (18–91 years, n = 351) of the sample available to Uttl than that in this study (65–103 years, n = 2,636). Perhaps of greater interest is the gradual decrease in NART performance with age. Age-related rates of decline remained after adjusting for basic sociodemographics and MMSE score, though they were greatly reduced. We did not adjust for health, mortality, or other dementia risk factors, which may further account for age-related declines in verbal abilities, as this was beyond the scope of this study. However, as previously mentioned, the possibility that true decline rates were masked by practice effects cannot be discounted. Thus it is important to note that despite the strongly supported notion of age stability for putative crystallized abilities in late adulthood (Baltes, Staudinger, & Lindenberger, 1999), subtle declines in verbal abilities, as indexed by the NART, may be expected in the general population.

The issue of cohort differences across studies is intriguing as it does not arise in typical reports of normative data that rely on single study samples. Even after adjusting for sociodemographic and health variables, the Adelaide sample on average scored lower than the Sydney and Canberra samples. This could be due to respondent characteristics not accounted for in the analyses, or study differences in administration protocols. Regardless of the underlying cause of these study differences in NART performance, this finding demonstrates a further advantage of analyzing harmonized and pooled data. The representativeness of the broader national population is enhanced.
while study-specific sample biases, which would otherwise be undetected, are minimized.

Limitations

Although imputation by ordinary least squares regression underestimated the variability in baseline NART scores for SOPS participants, the comparable distributions and strong association between the Schonell and NART indicate that this was of little consequence. In this context, little benefit was to be gained by predicting NART scores via more computationally and analytically complex missing data methods such as multiple imputation.

Responses to individual items were not available for the NART and Schonell in two of the three studies, which only coded summed total scores. Absence of individual item responses prevents reliability analyses, investigation of item characteristics, and the use of more robust techniques of harmonization such as those orientated by item response theory. This is an important reminder of the importance of having access to raw item-level data.

A possible shortcoming of this study is that these NART norms do not include participants from a nonurban population. In addition, we do not report NART norms for adults under the age of 65 years and acknowledge the small cell counts for adults aged over 95 years.

In summary, this study successfully demonstrates the benefits afforded by the harmonization of functionally equivalent measures and reports NART norms for adults aged 65 years and older that have increased precision and representativeness than those previously available. These norms for adults under the age of 65 years and acknowledge the small cell counts for adults aged over 95 years.

In conclusion, the NART as an index of prior intellectual functioning is recommended for use in research and clinical practice. The NART is a reliable and valid measure that can be used to assess cognitive function in elderly populations. The use of the NART in combination with other cognitive assessments can provide a more comprehensive understanding of cognitive function in older adults.

REFERENCES


StataCorp. (2007). *Stata Statistical Software: Release 10*. College Station, TX: StataCorp.


