What's age got to do with reading?

Kevin Wheldall

Molly de Lemos

Craig Wright

Expressing reading ability in the form of a 'reading age' is a common practice within schools and in research on reading. Reading age is, perhaps, particularly attractive by virtue of its simplicity: when compared to chronological age it appears to indicate how far behind or in front of the 'expected' level the student is reading, it allows quick comparison of the reading levels of multiple students, and allows teachers to understand how much correction has to be made to the curriculum for students who have a delay in reading age. This article will show that, despite its apparent attractiveness, the concept of reading age is fundamentally flawed. (Some of these issues have been raised before in the past (Alexander & Martin, 2004; de Lemos, 2000; McNab, 2007; Wheldall & Beaman, 2000) but very little attention has been paid to their implications.)

How reading ages are constructed

First, let us look briefly at how reading ages are typically constructed. The developers of a new test will seek to obtain performance data on their new measure from an ideally large and representative sample of students across the age range that the test aims to cover. The sample is divided into a series of age groups, usually covering a range of about 3 months to 6 months, depending on the age range covered by the test (e.g., 7:0 years – 7:2 years, 7:3 – 7:5, and so on). The average raw score of each of these age groups is calculated, and this average score is converted to a 'reading age' based on the mid-point of the chronological age of the age norm group. For example, if the age range of the age norm group is 7:0 to 7:2, and the mean raw score of the age norm group is 48, then a raw score of 48 would equate to a reading age of 7:1. The same procedure is used for all the age groups in the standardisation sample. The raw scores are then plotted against age, with age (the midpoint of each age group) on the horizontal axis and the average raw score of each age group on the vertical axis. A smooth line is then drawn linking these points. The raw score corresponding to each age in terms of years and months can then be estimated from this smoothed graph. Note that the number of years of instruction the children have received is not taken into account in the construction of reading ages.

Problems with reading age – variability

So, what's wrong with reading ages obtained in this way? First, there is the problem of variability of performance and hence in raw scores for each age group. The reading age is based on the average score for the age group but some students will read better and score higher and some will read worse and score lower. For example, some children in a typical Year 4 class will score at a level more typical of Year 1 or Year 2 students while others will score at a level more typical of Year 5 and Year 6 students.

The latest results of the National Assessment Program, Literacy and Numeracy (NAPLAN) for reading in 2016 for Year 3 and Year 5 illustrate this point (ACARA, 2016). The average scaled score for Year 3 students was 425.6 (standard deviation = 85.6). For Year 5 it was 501 (standard deviation = 77.1). From our knowledge of the normal distribution ('the bell curve'), it can be estimated that roughly 20% of the Year 3 students scored at the average level or better for Year 5, whereas about 20% of Year 5 students scored at the average level for Year 3 or worse. In other words, the variability in reading performance for students within grade (year) is very large indeed, and tends to increase with age. (An illustration of the variability in the spread of scores with age and the overlap of scores at adjacent age levels is provided by McNab (2007), who showed the expected distribution of scores at each age level as a series of overlapping bell curves based on a normal distribution.)

Problems with reading age – different meaning at different ages

A second problem with reading ages is that the significance of a discrepancy between chronological and reading age changes depending on the age of the student. Take data from the Neale Analysis of Reading Ability 3rd Edition (NARA III; Neale, 1999), a test that until recently was widely used in Australia, as an example. A student aged 7:6 halfway through Year 2 whose reading age is 18 months below her chronological age is a very poor reader indeed. The student's score on the Accuracy part of the Neale is equivalent to ~1% of their grade peers. That is, better than only 1% of students in Year 2. In contrast, a student aged 12:6 in Year 7 whose reading age is 18 months below his chronological age is actually an average reader (equivalent to 27% of his peers).

Problems with reading age – reading is related to years of instruction, not age

The concept and method of determining reading age depends upon the assumption that age within grade is an important determinant of reading ability. It is certainly true that reading performance increases with grade level. However, older students within a given grade are, on average, not better readers than the younger students in the same grade.

Across the primary school years, reading performance correlations appear to be as strong or stronger with grade or year level (i.e., years of instruction) than with chronological age. Further, correlations between measures of reading performance and chronological age within grade tend to be small and insignificant.

Some years ago now, we looked at the results for reading from the Basic Skills Test (BST) in New South Wales that preceded NAPLAN. The BST used to be administered to all primary school students in state schools and to many students in the Catholic and independent sector schools in Years 3 and 5 in August of each year. The literacy component tested students' understanding of a range of written texts used in the primary key learning areas. Actual chronological age of the child was not collected as part of the BST testing regime and so calculation of correlations between age and BST score was not possible. However, students taking the test were required to indicate on the test protocol whether for Year 3 they were aged under eight years (and very few were), aged over eight up to nine years, or aged over nine years. Similarly, Year 5 students had to indicate whether they were aged under 10 years (again very few were), aged over 10 up to 11 years, or aged over 11 years. Given that the BST was administered to almost all students in Year 3 and Year 5 in the state, the numbers in these samples are very large and approach, in effect, population parameters.

If literacy performance is correlated with age within grade then we would expect to observe appreciably higher BST

mean scores in the older age group than in the lower age group within each grade. This was not the case for any literacy measure for Year 3 or Year 5 in any of the three years studied (1998 to 2000). For example, consider the means for Year 5 students for reading in 2000. The mean score for the 42,254 10-year-olds was 56.6 whereas the mean for the 17,314 11-year-olds was 55.8. This virtual population based study carried out over three successive years provided no evidence for any association between age and reading performance within grade.

By way of further illustration, BST performance data were collected in the context of a study relating the Wheldall Assessment of Reading Passages (WARP; Wheldall & Madelaine, 2013) with the BST (Madelaine & Wheldall, 2002). Chronological age data were available for a sample comprising 65 Year 3 students and 58 Year 5 students. Moreover, this sample of students was shown to be highly representative of the state population as a whole in terms of BST performance. (The average scores for the school on BST literacy were shown to be very similar to State averages at both Year 3 and Year 5 levels consistently over several years.) For Year 3 students the BST literacy measure was shown to correlate with age at 0.16, whereas for Year 5 students it correlated at 0.15. The correlations between the WARP and chronological age for these two samples were 0.07 for Year 3 and 0.26 for Year 5: no more than 7% of the variability in scores, at best, was attributable to age.

In sum, there is little or no relationship between age and reading performance within grade. Correlations with age across grades are the result of increasing years of instruction, not maturation. While learning to talk is a developmental process, reading is not. Reading performance is largely a function of the amount and quality of instruction received. Given that this is the case, it probably makes more sense to relate reading performance to years of instruction received rather than to chronological age when comparing children regarding their reading ability.

Examples of how reading age is (mis)used

Two students in a Year 4 classroom, Steve (age nine) and Mark (age 10) are both tested as having the same reading age of 9:6. We would commonly claim that Steve is six months ahead in reading while Mark is six months behind, and that the two students are a year apart in terms of reading performance. Yet they are both in the same class, they have both experienced the same amount (four to five years) of reading instruction, and they are both reading at the same absolute level as measured by the raw score of the reading test (given that reading age is simply a reflection of raw score). Why would we expect them to be performing differently just because they differ in chronological age?

Here is another example. Jenny in Year 4 is 9:0 but has a reading age of 8:6. Sarah is aged 10:0 but has a reading age of only 8:0. Being 'only six months behind', Jenny would still typically be regarded as being within the average range of performance for her age. She is unlikely to be seen as a cause for particular concern. But Sarah is perceived as two years behind what we would expect for her age and would therefore typically be considered to be (by definition) a low-progress reader and a very real cause for concern. Yet they are both in the same year at school, have experienced the same amount of reading instruction over the past five years, and are only a few points different in terms of level of absolute performance as indicated by raw score on the reading test.

Link between age-based reading ages and grade-based stanine scores

The extent to which relatively large differences in reading age can still be within the 'average' range of scores according to the expected normal distribution of scores can be illustrated by looking at the range of reading ages that fall within stanines 4, 5 and 6 on the 9-point stanine scale. These stanines correspond to standardised scores ranging from 89 to 110, in which 54 per cent of scores would normally be expected to fall. The norms for the NARA III (Neale, 1999) provide both reading ages, based on age norm groups, and percentile and stanine scores, based on norms for 'years of schooling'. From the norm tables for this test the reading ages corresponding to each stanine level for each year of schooling can be identified.

The table below provides a summary of the range of reading ages on the Reading Comprehension measure of the NARA III that fall within stanine levels 4 to 6, which marks the average range of scores expected at each year of schooling. This table indicates that in the first year of schooling the differences in reading age that fall within the average expected for this level is less than one year (10 months), but by the fifth year of schooling the differences in reading age that fall within the average expected for this level is just over four years (four years and three months). Thus, as the variability of scores on measures of reading comprehension increase with age, it can be expected that a range

Correspondence between reading ages and stanine scores on the NARA III (Neale, 1999) for students in their first to fifth year of schooling

	First year of school	Second year of school	Third year of school	Fourth year of school	Fifth year of school
Mean age	6:2	7:2	8:2	9:2	10:2
Range in reading age for stanines 4 to 6	6:0 – 6:9	6:3 – 8:3	7:3 – 9:5	8:0 – 11:9	8:5 – 12:7
Range of average scores in years and months	0:10	2:01	2:03	3:10	4:03

of up to four years in reading age can be expected as normal and within the average range of scores expected at older age levels.

False negatives in screening for early reading difficulties

Finally, reading age can be responsible for the identification of false negatives in screening for early reading difficulties; that is, identifying children as average readers when they are actually poor, low progress readers. Take the NARA III (Neale, 1999) as an example. A Year 2 student of seven years of age who is in their third year of school and who has a reading age of 7:0 on the Accuracy part of the test can actually be a poor reader. The student's reading accuracy is better than just 18% of Year 2 students. Yet in using reading ages the examiner/teacher might assume that the student is exactly where they would be expected to be given their age. The obvious problem with this is that the child fails to receive the intervention that is crucial for overcoming written language deficits.

A general comment on age norms versus grade norms

Reading tests tend to be constructed by assessing all students at one point in the school year. Norms are then generated for different age groups by pooling the data for all students in a given age range (e.g., 7:0-7:3, 7:4-7:7 and so on).

This practice causes two problems for test users. First, students in the same age range may actually be in different grades (and we have shown above that reading ability is related to years of instruction rather than to age within grade). This issue potentially makes interpreting age-based norms very problematic. Second, normative data is typically only collected at a single time within the school year. For example, the manual for the NARA III (Neale, 1999) states, "The standardisation took place from September to November 1997 during the final term of the Australian school year". The time at which the normative data were collected can have a big effect on interpretation because reading ability changes so much over the course of a school year; particularly in younger grades. A test for which data are collected in Term 4 is likely to under-estimate the skills of a student tested in term 1 of that year. This problem has led Vincent (1997) to rightly argue that test norms "will only accurately reflect children's attainments at the time of year at which they were obtained. This seemingly obvious point is too often overlooked by test users …"

We suggest that the gold standard for tests should be data collected in each of the four Australian school terms.

Note that these latter concerns apply to all normative scores obtained from tests, not just reading ages. Standardised scores, z-scores, percentile ranks and stanines all suffer from the same criticisms regardless of whether they are normed on the basis of age or on the basis of grade level.

The solution? Create tests that are standardised at separate time intervals over the school year. We suggest that the gold standard for tests should be data collected in each of the four Australian school terms. A less acceptable alternative would be data collected in the two Australian semesters; preferably at the mid-point of each semester to minimise false positives and false negatives at the beginning and end of each semester period respectively. (Some test developers have begun to take this

problem on board. The Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner & Rashotte, 2012) and the Wechsler Individual Achievement Test (WIAT-III; Psychological Corporation, 2016) both provide grade-based norms for two time points in the academic year.)

Given the increasing difficulty and expense incurred by test publishers in providing norms for reading and other performance tests, perhaps we should not hold our breath. In the meantime, we suggest that test users think critically about the quality of normative data available for any given test before purchasing or using the test. We also urge test users to interrogate the scores obtained from any test by considering how representative the normative data are for the student in question (e.g., by considering the time of year at which the data were collected, the number of students in the sample and whether number of years of instruction has been accounted for) before making conclusions and high-stakes decisions. For researchers and clinicians seeking to measure progress across time, we suggest using raw scores rather than standardised scores or reading ages.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA; 2016). *NAPLAN Achievement in Reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2016*. Sydney: Author. Retrieved from https://www.nap.edu.au/results-and-reports/national-reports

Alexander, J. R. M., & Martin, F. (2004). The end of the reading age: grade and age effects in early schooling. *Journal of School Psychology, 42*, 403-416. doi:10.1016/j.jsp.2004.06.003

de Lemos, M. M. (2000). *Group reading test II: Australian norms supplement*. Camberwell: ACER Press.

Madelaine, A., & Wheldall, K. (2002). Further progress towards a standardised curriculum-based measure of reading: Calibrating a new passage reading test against the New South Wales Basic Skills Test. *Educational Psychology, 22*, 461-470. oi:10.1080/0144341022000003132

McNab, I. (2007). Making sense of reading ages. *LDA Bulletin of Learning Difficulties Australia, 39*(2&3), 4-5. Retrieved from https://www.ldaustralia.org/lda-publications.html#bulletin

Neale, M. D. (1999). *Neale Analysis of Reading Ability (3rd ed.)*. Melbourne: Australian Council for Educational Research.

Torgesen J. K., Wagner R. K., & Rashotte C. A. (2012). *Test of Word Reading Efficiency (2nd ed.)*. Austin, TX: Pro-Ed.

Vincent, D. (1997). Assessment by classroom teachers. In R. Beech & C. Singleton (Eds.), *The psychological assessment of reading* (pp. 27-48). London: Routledge.

Psychological Corporation (2016). *Wechsler Individual Achievement Test (WIAT; 3rd ed.)*. Sydney: Pearson.

Wheldall, K., & Beaman, R. (2000). *An evaluation of MULTILIT: 'Making Up Lost Time In Literacy'*. Canberra: Department of Education, Training and Youth Affairs.

Wheldall, K., & Madelaine, A. (2013). *The Wheldall assessment of reading passages (WARP) manual*. Sydney: MultiLit Pty Ltd.

Emeritus Professor Kevin Wheldall AM is Chairman of MultiLit Pty Ltd and Director of the MultiLit Research Unit. You can follow him on Twitter (@KevinWheldall) where he comments on reading and education (and anything else that takes his fancy). He also has a blog, 'Notes from Harefield: Reflections by Kevin Wheldall on reading, books, education, family, and life in general': www.kevinwheldall.com. Email: kevin.wheldall@pecas.com.au

Dr Molly de Lemos AM is a psychologist and former educational researcher with the Australian Council of Educational Research. While at ACER she spent some time on the development and norming of educational and psychological tests. She is the author of the report, Closing the Gap between research and practice: Foundations for the acquisition of literacy, *published by ACER, and is a former President of Learning Difficulties Australia (2013/2014). Email: delemos@pacific.net.au*

Dr Craig Wright is a psychologist and Clinic Director at Understanding Minds and Adjunct Research Fellow in the School of Psychology at Griffith University. He avoids social media but can occasionally be found blogging at www.understandingminds.com.au/blog where he writes about issues relating to developmental and learning disorders. Email: craig@understandingminds.com.au