Precision, Reliability and Application of the Wilkins Rate of Reading Test

James Gilchrist[1], Peter M. Allen[2], Laura Monger[2], Krithica Srinivasan[3], Arnold Wilkins[4]

1. Independent Researcher in Optometry and Vision Science, North Yorkshire, United Kingdom
2. Vision and Hearing Research Institute, Anglia Ruskin University, Cambridge, United Kingdom
3. Department of Optometry, Manipal College of Health Professions, Manipal Academy of Higher Education, Manipal, India
4. Department of Psychology, University of Essex, Colchester, United Kingdom

Running head: Wilkins Rate of Reading Test

Keywords: precision, reliability, fluency, reading age

Address for correspondence

Dr James Gilchrist
j.m.gilchrist@gmail.com

Abstract

Background: The Wilkins Rate of Reading Test (WRRT) enables rapid measurement of reading speed using text passages that have no semantic content and impose minimal demand on word recognition (decoding) skills. It is therefore suited to applications where the primary interest is in the influence of visual and ocular motor factors on the rate of reading. The test has been used to assess the effects on reading of (1) type design; (2) binocular function; (3) 2D and 3D displays and (4) coloured filters.

Methods: We obtained estimates of the precision and reliability of WRRT using test and retest measures of reading rate from four samples collected independently by the authors: JG n = 120 adults; PA & LM n = 100 adults; KS n = 799 children; AW n = 139 children. Each participant was asked to read aloud as quickly and accurately as possible, for one minute, and results were recorded as number of words read correctly per minute (wcpm).

Results: Precision estimates from each sample are given by the within-subject standard deviation s_w and reliability estimates by the intraclass correlation coefficient r_i. and for each sample the precision and reliability estimates are as follows: JG s_w = 11.4 wcpm, r_i = 0.87; PA & LM s_w = 5.4 wcpm, r_i = 0.97; KS s_w = 7.9 wcpm, r_i = 0.91; AW s_w = 7.4 wcpm, r_i = 0.92.

Conclusion: The reliability of WRRT is high (>0.85) in all four samples, reflecting large variation in reading rate between subjects compared to within-subject variability. This indicates that WRRT is an excellent test for discriminating differences in reading speed between individuals. The precision of the test varies from 5.4 to 11.4 wcpm among samples, with a pooled value of 8.3 wcpm, providing a basis for setting criteria for change. Based on these data, we propose that interventions aimed at improving reading rate should result in an increase of at least 17 wcpm (2 x s_w), or 25 wcpm (3 x s_w) for greater confidence that the improvement exceeds test-retest variation.

The goal of reading is comprehension of text, which requires both the ability to identify words and fluency in doing so.[1,2] Word identification requires mapping of word orthography to phonology (print-to-sound decoding), recognition of whether the result constitutes a word and a decision concerning its meaning.[3,4]

Fluency involves the ability to read words quickly with natural intonation and expression (prosody), and it is regarded as a key link between word identification and comprehension.[5] Comprehension requires fast and efficient execution of the word identification process, and the ability to render the resulting stream of words with sufficient speed.[5]

Whereas single word identification may be assumed to depend primarily on orthographic-phonological decoding, prosody depends on the fluency with which a word sequence can be read. Fluency is strongly affected by certain visual and ocular motor factors.[6] The influence of these factors can be measured separately from the cognitive factors that underpin decoding. Decoding ability alone can be measured by the accuracy of single word and non-word identification, without respect to speed. Comprehension of text can be measured without regard to accuracy of specific word identification or reading rate, and fluency can be measured as the rate of reading sequential text. Ideally fluency is measured when the decoding and comprehension demands are minimised, because decoding and comprehension abilities themselves may help or hinder reading rate.[3]

Wilkins Rate of Reading Test

The Wilkins Rate of Reading Test (WRRT) was introduced to provide a test that could be used to evaluate the effects on reading speed of visual factors and interventions (notably lenses and/or filters), especially in children with reading difficulties.[7] The design of the test “minimises the linguistic and semantic aspects of reading and maximises the visual difficulties”, noting that “many visual difficulties with reading seem to emerge when the test is presented in a long paragraph with closely spaced lines and letters.” [7] The test is not a test of fluency, at least in so far as fluency includes normal prosody, because prosody is altered when words are disconnected and the test meaningless, as in this test.

The text of the WRRT (Figure 1) is reproduced in a small (9pt) self-similar font (Times New Roman) with a small (4 pt) space between words. The text is set as a paragraph of ten lines 72.5 mm wide, 33.4mm high, with an interline space of 3.15mm. The letters have an x-height of 1.6 mm and a width that averages 1.53 mm. The text consists of 15 high-frequency words, the same 15 words on each of the ten lines, but in a different random order. Although one word in the passage may cue another neighbouring word with which it is commonly associated (e.g. cat-dog), this association is random and will be similar overall from one version of the test to another. The test can be tackled both by adults and by children who have only a modest reading vocabulary.

The WRRT is unlike other tests of reading in that it is not primarily designed to compare one individual with others of the same age or ability but rather to measure the effects of interventions that affect reading rate in specific individuals. Although the WRRT has been in existence for more than 20 years, there are only limited data on its precision and reliability. Beyond its use by optometrists and others for assessments related to visual stress,[8] the test remains largely unknown, and its potential for more general use as a measure of reading ability unrecognised outside this area of application. Our aim here is to address these issues. We first present data that provide estimates of the precision and reliability of WRRT in schoolchildren and young adults, and later we consider its application in a variety of contexts.

Figure 1. The Rate of Reading Test (from Wilkins et al., 1996)

come see the play look up is cat not my and dog for you to
the cat up dog and is play come you see for not to look my
you for the and not see my play come is look dog cat to up
dog to you and play cat up is my not come for the look see
play come see cat not look dog is my up the for to and you
to not cat for look is my and up come play you see the dog
my play see to for you is the look up cat not dog come and
look to for my come play the dog see you not cat up and is
up come look for the not dog cat you to see is and my play
is you dog for not cat my look come and up to play see the

see the look dog and not is you come up to my for cat play
not up play my is dog you come look for see and to the cat
look up come and is my cat not dog you see for to play the
my you is look the dog play see not come and to cat for up
for the to and you cat is look up my not dog play see come
you look see and play to the is cat not come for my up dog
come not to play look the and dog see is cat up you for my
and is for dog come see the cat up look you play my not to
dog you cat to and play for not come up the see look my is
the come to up cat my see dog you not look is play and for

Concepts of Precision and Reliability

Measures of human performance or ability can serve two purposes: i) to enable monitoring of changes in ability within individuals over time, and ii) to reveal differences in ability between individuals. In relation to reading ability, measurement of change within-individuals is of particular importance for monitoring the development of reading ability in children and the deterioration of reading ability in older adults experiencing loss of function such as visual or cognitive impairment. In both cases, repeated measurements over a period of time serve not only to reveal the pattern of change but also to demonstrate whether interventions are of benefit in helping to improve development of an individual's reading ability or slow its decline. On the other hand, measuring differences between individuals enables identification (diagnosis) of those whose reading ability is substantially lower than their peers, and provides evidence to support the introduction of interventions aimed at improvement. The statistical requirements of an effective test for the two purposes just described are, respectively, *precision* and *reliability*.

Precision is the general term for variability between repeated measurements from the same individual and takes account only of the variation *within*-subjects, not that *between*-subjects. In order to be effective for monitoring change within-individuals over time, a measure should be precise; that is, the test-retest variation due to measurement error should be stable and low. The terms precision, repeatability and reproducibility all relate to within-subject variation, that is, the consistency of repeated measurements.[9] Reliability, on the other hand, takes account of variation both within- and between-subjects. Reliability is the degree to which variation between-subjects exceeds that within-subjects.[10] When between-subjects variation is large compared to that within-subjects, then test reliability is high and scores from different individuals are likely to indicate real difference between them rather than the effects of measurement error. Conversely when between-subjects variation does not greatly exceed that within-subjects, then test reliability is low and score differences between individuals may reflect the effects of measurement error rather than true differences. Reliability may be thought of as a measure of *discriminability*, as it emphasises the principle of being able to use the test to discriminate reliably between different individuals.

Finally, the recognition that precision is a measure of variability within-individuals, while reliability involves both within- and between-individual variability, means that we can expect two different scenarios in practice. The first relates to tests that are intended to be used in a single population in which the variation between-subjects is assumed to be constant. In this case, the better the precision of a test, the better will be its reliability; in other words, reliability will be determined by precision. The second scenario relates to tests employed in a number of different populations, each with its own degree of between-subjects variation. In this case, it is quite possible for a test having good precision to have good reliability in one population and poor reliability in another. This highlights the importance of making a clear distinction between the concepts of precision and reliability, and the need for test evaluation to be undertaken in different populations as appropriate.

We will now apply these concepts to data obtained using the WRRT in four samples of participants drawn from different populations.

Methods

Participants and Data Collection

Participants were recruited and assessed by the authors in their respective university locations, giving four separate samples having the characteristics summarised in Table 1.

Table 1. Samples used to assess precision and reliability of the WRRT

Sample	Author/s	Sample size	Sample characteristics
A	Gilchrist	120	Adults: Bradford, UK. mean age 21.8 yrs, range 18-40 yrs, 57% female
B	Monger & Allen	100	Adults: Cambridge, UK. mean age 21.4 yrs, range 17-31 yrs, 63% female
C	Srinivasan	799	Children: Udupi Taluk, India mean age 11.7 yrs, range 7-16 yrs, 46% female
D	Wilkins	139	Children: Norwich, UK mean age 10.5 yrs, range 9-12 yrs, 59% female

Adult participants in Samples A and B were recruited from undergraduate student populations attending the University of Bradford and Anglia Ruskin University, Cambridge respectively. Children in Sample C were recruited from 9 different schools located in the Udupi Taluk region of India, while those in Sample D were recruited from 1 school in Norwich, UK. Adults in Samples A and B were tested in their respective universities, while children in Samples C and D were tested in the schools they attended. Data for Sample A were collected by two undergraduate student assistants who had been trained in the use of the WRRT but were not involved in study design, data analysis or authoring, and each participant was assessed by only one of the student assistants. Data for Samples B and C were collected by the authors LM and KS respectively, while those for Sample D were obtained as part of a study that has been published previously.[11] Recruitment and participation of subjects was achieved in compliance with relevant local/institutional requirements for ethical approval. In the case of the children in Samples C and D, parental consent for participation was obtained.

All participants provided two measures of reading rate with the WRRT. These test and retest measures provide all the data to be presented in the analysis that follows below. In each of Samples B to D, the collection of WRRT measurements was undertaken as part of investigations of the effects of coloured overlays on reading rate.[11] For these samples, therefore, the WRRT data to be presented here are those taken without the use of any coloured overlay but obtained within a testing sequence that interleaved WRRT measurements with and without coloured overlays. In the case of Samples B and D, the WRRT test sequence in relation to use of coloured overlays was *with-without-without-with*, while for Sample C the WRRT test sequence involving overlays was allocated to subjects randomly as either *with-without-without-with* or *without-with-with-without*. The tests were presented in immediate succession, typically less than 1 minute apart. For Sample A there was no use of coloured overlays at any stage of the data collection.

Participants in all samples, including the children in Sample C whose native language was not English, demonstrated that they were able to read (recognise and pronounce) the 15 words included in the WRRT prior to testing. The test was scored by noting the errors on a score sheet comprising an enlarged version of the text, and by measuring the total time taken to read the passage. From these measurements the reading rates were estimated as words correct per minute (wcpm). Participants in Sample D read the entire passage, while those in the three other samples read for one minute. The passage length in WRRT (150 words) is such that the time difference between reading the whole passage and reading for one minute is typically small.

All subjects who would normally use refractive correction for their academic or schoolwork were corrected for this study, otherwise no refractive correction was given. The test conditions for all samples were controlled by fixing the viewing distance at ~40cm, with lighting conditions adjusted to give a glare-free illuminance level on the task of 500 to 1000 lux, resulting in task background luminance between 70 and 100 cd/m^2. For samples A and B the lighting was a tungsten-halogen desk lamp adjusted to give the luminance described above, with ambient room lighting provided by a ceiling-mounted 'warm white' fluorescent lamps. For sample C, natural daylight was available, while for Sample D the lighting was fluorescent, with magnetic ballast. For all samples, selection of which of the four standard WRRT passages to use was randomised and every subject was presented with a different passage on test and retest (see Figure 1).

Statistical Analysis

As described in the introduction above, the principal aim of data analysis was to obtain estimates of WRRT precision and reliability. The statistical principles are set out in some short notes by Bland & Altman[12] and elaborated in greater detail in several texts.[10,13,14]

Precision

The underlying assumption, consistent with classical test theory,[10] is that each subject has a *true* reading rate, which is estimated by each individual measurement. The best estimate of true reading rate will be the average of a number of repeated measurements and, assuming that the true rate remains constant (at least over the relatively short time periods that apply here), then repeated measurements of the same subject may be assumed to vary around the true value because of measurement error. Thus, the standard deviation of repeated measurements on any individual subject will provide an estimate of measurement error. If such an estimate is obtained from a number of participants then its value will vary, so to estimate the measurement error in a sample involves calculating the common within-subjects standard deviation in the sample.[12] This value is the estimated precision of the test, denoted s_w. Note that this means that the precision of a test is expressed in the units of measurement of the test. Small values of s_w indicate small amounts of measurement error, which corresponds to good precision. An important caveat for the use of s_w as an overall estimate of measurement error in a sample of subjects is that there should be no evidence that the size of s_w is systematically related to the level of performance on the test. This can be checked by plotting the standard deviation of repeated measurements (or their absolute difference, in the case of test and retest measures) against their mean.[12]

Reliability

The use of the correlation between test and retest scores as a measure of the reliability of the test is discussed by Bland & Altman[12] who recognise its interpretation not as a measure of the amount of measurement error but of the ability of a measure to discriminate between individuals: "The correlation coefficient [between one test and the next across participants] can be used to compare measurements of different quantities … The measures with the highest correlation between repeated measurements would discriminate best between individuals." Bland and Altman[15] point out that the correct approach to this is to use the intraclass correlation coefficient, which estimates the degree to which the variation in measurements between-subjects exceeds that within-subjects. Note that the reliability of a test is expressed by a value between 0 and 1 with no measurement units. This is consistent with the interpretation of reliability as a correlation between test and retest measurements.

Calculations

We calculated estimates of precision and reliability using a one-factor, repeated-measures analysis of variance (ANOVA), in which the repeated measures were test and retest reading rates for each participant. This approach is endorsed by Bland and Altman,[12,15] and full details of the ANOVA model and the necessary computations are given.[14] The foundation of the analysis is the principle that the total variation in the data can be portioned into between-subjects and within-subjects components, which ANOVA expresses as measures of mean-squared variation. The estimated precision of a single reading rate measurement is given by the square root of the within-subject mean square, i.e. $s_w = \sqrt{MS_w}$. The reliability of a single reading rate measurement, estimated from two repeated measurements (test & retest) on each participant, is given by the intraclass correlation coefficient, $r_i = (MS_b - MS_w)/(MS_b + MS_w)$, where MS_b is the mean square variation between-subjects.[14]

Results

Descriptive Statistics and Comparisons

Figure 2 shows that distributions of the reading rates (wcpm) are approximately normally distributed in all four samples of participants, and summary statistics are given in Table 2.

Figure 2. Reading Rate distributions

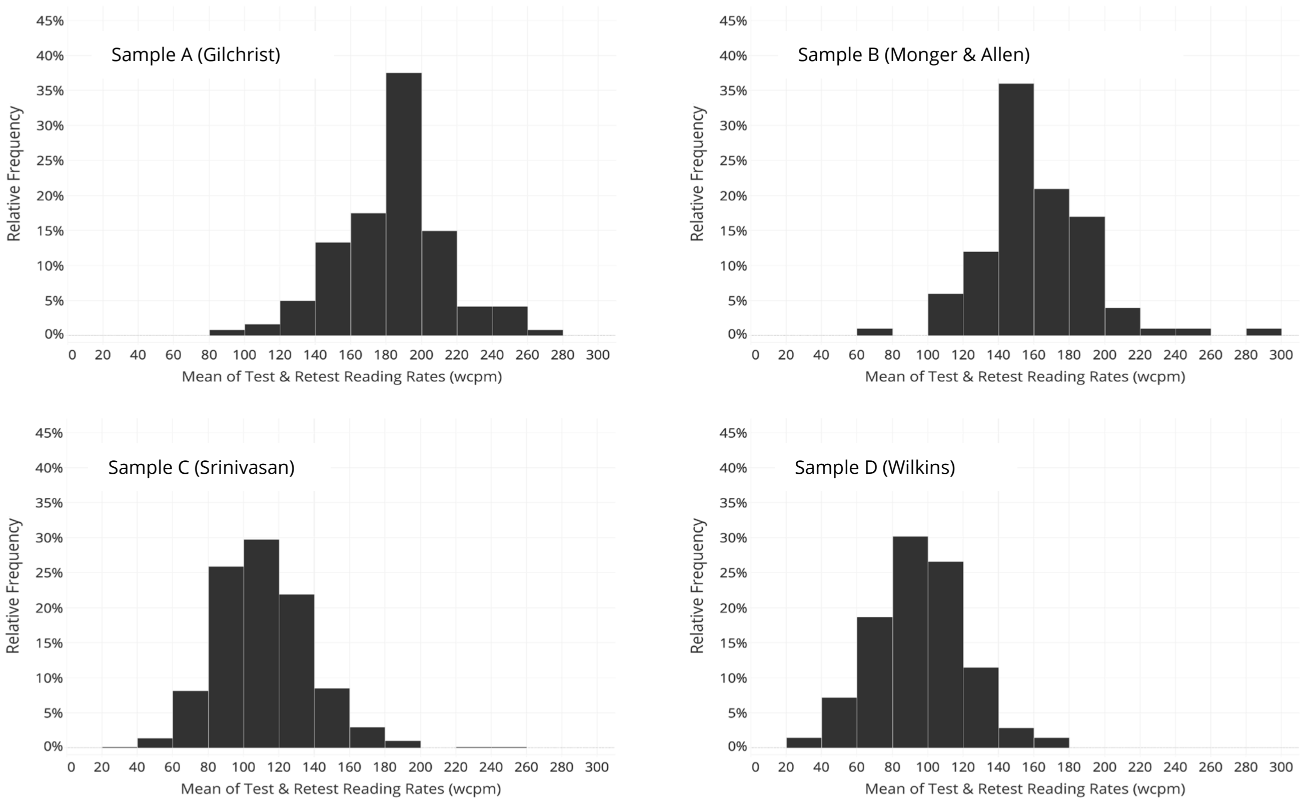

Table 2. Summary Statistics of Reading Rate distributions

Sample	n	Mean	SD	Min-Max (Range)	Median	Skewness	Kurtosis	Shapiro-Wilk W	Shapiro-Wilk p
A	120	183.5	30.2	95 - 267 (172)	187	-0.061	0.540	0.985	0.205
B	100	160.7	29.3	73 - 283 (210)	156	0.746	3.040	0.955	0.002
C	799	111.4	26.4	40 - 240 (200)	110	0.584	1.100	0.982	<0.001
D	139	95.4	25.9	32 - 168 (136)	96.5	0.108	0.088	0.994	0.825

Note that standard deviations of the four samples are similar and their values indicate a general finding of large variation in reading rates between participants, with an overall average range of 180 wcpm between slowest and fastest readers.

Mean reading rates differed significantly among the four samples in the study: $F(3{,}1154) = 360.2, p < 0.001, \eta_p^2 = 0.484$, and as expected the rates for adult participants (Samples A and B) are higher than those for children (Samples C and D): $\mathrm{F}(1{,}1156)=936.3$, $\mathrm{p}<0.001$, $\eta_\mathrm{p}^2=0.448$.

Comparison of test and retest values in each sample (Table 3) shows no evidence of difference across repeated measurements. This suggests that reading rates overall were stable, and that the test showed little in the way of practice or fatigue effects, despite the unusual nature of the task.

Table 3. Sample mean reading rates on repeated (test & retest) measurements

Sample	n	Mean1 'test'	SD1 'test'	Mean2 'retest'	SD2 'retest'	t	df	p
A	120	182.7	29.7	184.3	32.8	1.050	119	0.298
B	100	160.4	29.5	160.9	29.5	0.613	99	0.541
C	799	111.7	27.2	111.1	26.8	1.570	798	0.117
D	139	96.1	25.9	94.7	26.9	1.650	138	0.102

Precision

As discussed previously, estimation of test precision using the overall within-subject standard deviation s_w assumes there is no association between participants' variability and their mean scores. In other words, the measurement error of the test should not depend upon whether a subject is a slow or a fast reader. Figure 3 shows the association between the absolute test-retest difference and mean reading rate for participants in each of the four samples in the study. Data from adult participants (Samples A and B) show no evidence of any association between variation of repeated measurements and their mean, indicating that the precision of WRRT in adults is independent of whether individuals are slow or fast readers. Data from children (Samples C and D) show a weak positive trend, with average test-retest differences increasing slightly as reading rate increases. Given the general lack of evidence of association between mean reading rate and magnitude of variation, we conclude that using the overall within-subject standard deviation s_w as an estimate of test precision in each sample is justified, but the implications of this will be discussed later. On this basis, Table 4 gives point estimates and confidence intervals for the precision of WRRT measurements in each sample.

Figure 3. Association of within-subjects standard deviation and mean reading rate

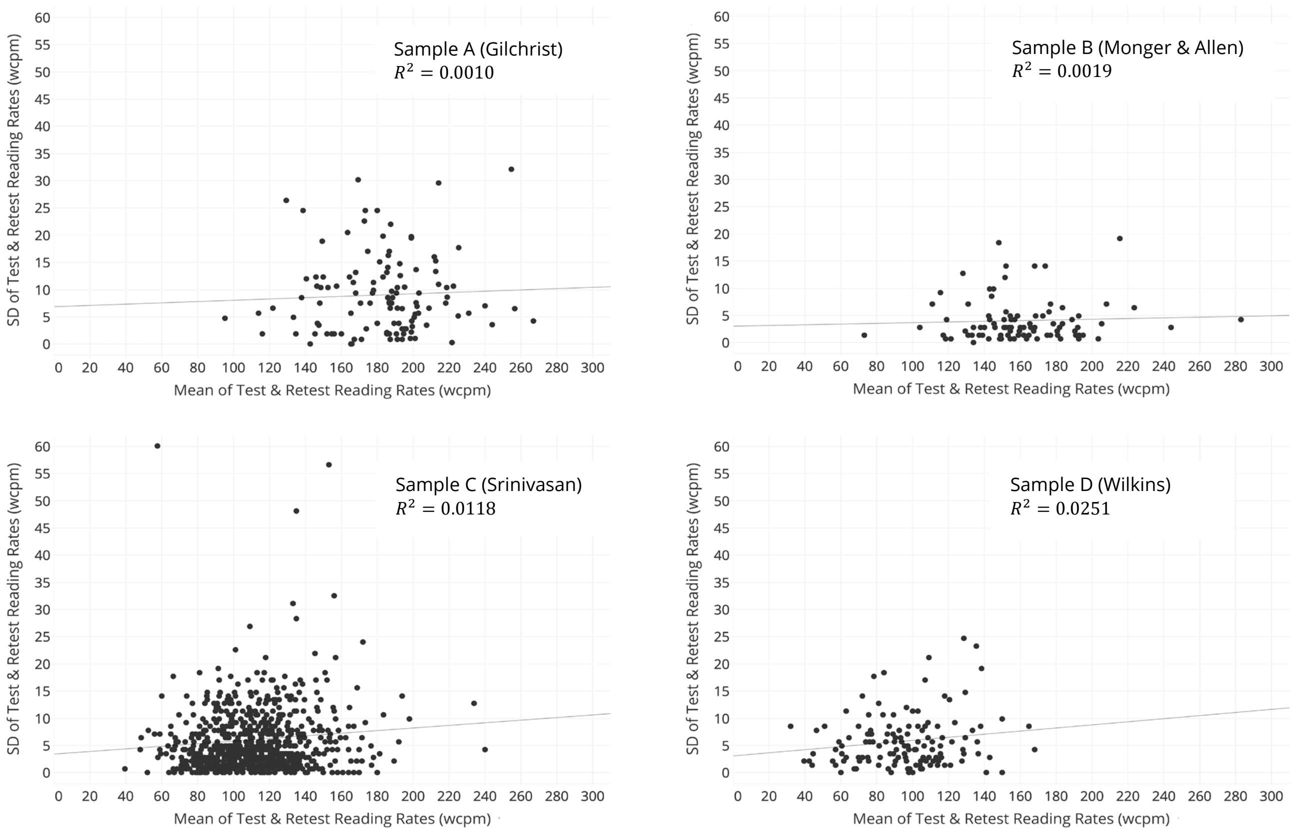

Table 4. Point estimates of WRRT precision, and 95% confidence limits

Sample	Precision $s_w = \hat{\sigma}_w$ (wcpm)	$\hat{\sigma}_w$ lower CL (wcpm)	$\hat{\sigma}_w$ upper CL (wcpm)	95% criterion for change (wcpm)	99% criterion for change (wcpm)
A	11.4	10.1	13.1	~23	~34
B	5.4	4.8	6.3	~11	~16
C	7.9	7.6	8.3	~16	~24
D	7.4	6.6	8.4	~15	~22

The general application of precision estimates is to set limits within which repeated measurements are expected to vary due to measurement error alone, so that a criterion can be set for the required effect of any intervention that purports to bring about a change in true reading rate. For example, to apply this principle to Sample A, we take the point estimate of $s_w = 11.4$ wcpm as the basis of the criterion for change and then, depending upon how strict the criterion needs to be, set a multiple of s_w (typically 2x or 3x for 95% and 99% limits respectively) as the value that must be exceeded. Thus, setting a criterion of $2 \times s_w \cong 23$ wcpm means that an intervention would need to change the reading rate of an individual by at least 23 wcpm to be confident that the change was greater than expected due to measurement error. The last two columns in Table 4 give the estimated change criteria for each sample in the study, based on these principles.

Reliability

As discussed previously, reliability is estimated using intraclass correlation coefficients, which may be calculated readily from mean-squares values given by ANOVA.[14] Table 5 shows the intraclass correlations of test 1 with test 2 for each sample, together with their 95% confidence limits.[16,17] Values of Pearson's and Spearman's correlation are included for comparison; although the intraclass correlation coefficient is the correct statistic to use in evaluating test-retest reliability, in practice we see that it gives values similar to those given by the more familiar correlation coefficients.

Table 5. Point estimates of WRRT reliability, and 95% confidence limits

Sample	Reliability $r_i = \hat{\rho}_i$ (intraclass correlation)	$\hat{\rho}_i$ lower CL	$\hat{\rho}_i$ upper CL	Pearson's correlation	Spearman's correlation
A	0.867	0.728	0.938	0.871	0.820
B	0.966	0.933	0.983	0.966	0.948
C	0.914	0.609	0.984	0.914	0.913
D	0.922	0.823	0.966	0.923	0.923

Discussion

Precision and Reliability of the WRRT

Precision estimates enable specification of a criterion for change within individuals. This is essential when WRRT is to be used for monitoring natural development or decline in reading rate, or the effect of some intervention. Sample values for precision in this study range from $s_w = 5.4$ to 11.4 wcpm, and the pooled $s_w = 8.3$, with corresponding population estimate (95% CI) of 8.00 <= $\hat{\sigma}_w$ <= 8.7. Based on this estimate, we might adopt a somewhat conservative 'universal' change criterion of $2 \times s_w$, where $s_w = 8.5$ (for ease of calculation), so that a significant change in reading rate would be taken as increase or decrease of 17 wcpm. Note that if this criterion of 17 wcpm is applied in our two samples (C & D) involving children, which combined have a mean reading rate of 109 wcpm, then it represents a change of approximately 15%, which has been recommended previously for use in practice.[18]

A caveat in the use of percentage change, however, is that the same change criterion (17 wcpm) applied to the adult samples (combined mean reading rate 173 wcpm) represents a change of approximately 10%, not 15%. Thus, we favour expressing the desired criterion for change in wcpm, rather than as a percentage of the initial value. The use of a criterion expressed as wcpm, rather than as percentage change, is justified by the finding, in all 4 samples, that the magnitude of within-subject reading rate variation is independent of mean rate. This is not to say that a change in reading rate cannot usefully be expressed as percentage change, only that in setting the criterion for when the magnitude of change exceeds the expected measurement error then we should apply this criterion in wcpm.

A very important caveat in the adoption of a single value as a criterion for change is clearly apparent in Figure 3, where we see not only that s_w varies between samples but also that it varies to a much greater extent between individuals within samples. The extent of this variation is such that, in each sample, there are individuals who exhibit test-retest variation much lower than the overall sample s_w, while other individuals exhibit much larger variation. For this reason, a strong case could be made that in every situation where monitoring of individual change is of concern then the change criterion should be based on an estimate of the baseline variation of that individual, rather than on a single, generalised estimate.

Reliability estimates evaluate the degree to which variation between-subjects exceeds that within-subjects, and thereby indicate the ability of a test to discriminate individual differences. Our results show that the WRRT has exceptionally high reliability in all the populations sampled, and so is an excellent test for the purpose of identifying differences in reading rate between individuals.

The high reliability of the WRRT reflects the fact that, in all populations but particularly in children, there is a very large variation in rate of reading between individuals (see Figure 2). The wide variation in reading rates between participants in our study is striking in all the samples we examined. Some children read faster than the average adult and, conversely, some adults read slower than the average child. It has previously been noted that within children who have similar scholastic attainment in reading, the variation in reading rate from one individual to another is more than a factor of 3, both in 7 year-olds[11] and 13 year-olds.[19] In the present study the variation in children's reading rate is even larger; whereas the adults (Samples A & B) showed an average of ~3.3x difference in reading rate between slowest and fastest, in children (Samples C & D) those with the highest reading rate were ~5.6x faster than those with the lowest.

Applications of the WRRT

There are many tests of reading ability, some aimed at children who are acquiring reading skills, and others aimed at skilled adult readers whose reading ability is somehow impaired, perhaps by dyslexia, loss of vision or cognitive decline. Some of these tests use passages of meaningful text to assess comprehension, word identification accuracy (decoding ability) and/or reading rate (fluency), while others use isolated words and sometimes non-words to assess decoding accuracy or efficiency (i.e. rate). Unlike the WRRT, however, none of these tests attempts to separate assessment of reading rate/fluency from assessment of decoding ability and comprehension, and the consequence is that the influences of cognitive and language skills that underpin decoding and comprehension are confounded with the influences of visual and ocular motor skills and speed of processing/naming, which are important in determining rate of reading.

Here we have shown that the statistical properties of the WRRT support its use for monitoring reading rate change within individuals over time, and also for assessing differences in reading rate between individuals. The WRRT can be used with people of any age, including young children having limited word knowledge and vocabulary and, by minimising or eliminating the influences of decoding ability and comprehension, the WRRT provides a measure of reading rate that should be more sensitive than other tests to the influences of visual and ocular motor factors.

Although the WRRT has been most widely used to assess the effects of a particular form of visual intervention (coloured overlays) on reading rates in children,[7,20] it has also been used in other contexts in which primary interest is the effect on reading of visual and/or ocular motor factors. For example, in previous studies, reading rate on the WRRT has been shown to be affected by aspects of typography such as the spatial periodicity of text,[21] font size (x-height) and font design in reading schemes for children.[21,22] Other researchers have used WRRT to assess the effects of treatment in cases of visual asthenopia[23] and binocular vision anomaly,[24] and to assess whether individuals using 3D displays may be susceptible to visual fatigue due to the demands of such displays on binocular visual and ocular motor functions.[25]

Finally, and more generally, we note that the WRRT is in effect a test of rapid automatized naming (RAN), in which the stimuli happen to be automatized words rather than letters, digits, etc., as is common in some applications of RAN. Although it has not previously been presented as such, considering the WRRT from this perspective greatly broadens its potential scope of application, as it is widely acknowledged that RAN performance is an important predictor of reading attainment.[26]

References

1. Verhoeven, J., & Perfetti, C. (Eds.) (2008). Advances in text comprehension: Model, process and development (Special Issue). *Applied Cognitive Psychology, 22* (3).
2. Garcia, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research,* 84, 74–111.
3. Hess, A. M. (1982). An analysis of the cognitive processes underlying problems in reading comprehension. *Journal of Reading Behavior*, XIV, 313–333.
4. Kendeou, P., van den Broek, P., Helder, A. & Karlsson, J. (2014) A cognitive view of reading comprehension: implications for reading difficulties*. Learning Disabilities Research and Practice* 29(1), 10-16.
5. Bashir, A.S., & Hook, P.E. (2009). Fluency: A key link between word identification and comprehension. *Language, Speech, and Hearing Services in Schools*, 40, 196–200. doi:10.1044/0161-1461(2008/08-0074).
6. Reichle, E.D., Pollatsek, A., Fisher, D.L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157. https://doi.org/10.1037/0033-295X.105.1.125
7. Wilkins, A.J., Jeanes, R.J., Pumfrey, P.D., & Laskier, M. (1996). Rate of Reading Test: its reliability, and its validity in the assessment of the effects of coloured overlays. *Ophthalmic and Physiological Optics*, 16, 491– 497.
8. Wilkins, A.J. (1995). *Visual stress*. Oxford: Oxford University Press.
9. ISO 5725-2:2019 Accuracy (trueness and precision) of measurement methods and results — Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. https://www.iso.org/standard/69419.html.
10. Allen, M.J., & Yen, W.M. (1979). Introduction to measurement theory. Monterey (CA): Brooks/Cole.
11. Scott, L., McWhinnie, H., Taylor, L., Stevenson, N., Irons, P., Lewis, L., Evans, M., Evans, B., & Wilkins, A. J. (2002) Coloured overlays in schools; orthoptic and optometric findings. *Ophthalmic and Physiological Optics*, 22, 156– 165.
12. Bland, J.M., & Altman, D.G. (1996)a. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307–10.
13. Dunn, G. (1989). Design and Analysis of Reliability Studies: The Statistical Evaluation of Measurement Errors. London, Edward Arnold.
14. Winer, B.J. (1971) Statistical Principles in Experimental Design McGraw-Hill, Kogakusha, Tokyo
15. Bland, J.M., & Altman, D.G. (1996)b. Measurement error. *BMJ*, 312 (7047), 1654.
16. Shrout, P.E., Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*, 86, 420– 428.
17. Bonett, D.G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med*, 21, 1331–1335.
18. Wilkins, A.J., Allen, P.M., Monger, L.J., Gilchrist, J.M. (2016). Visual stress and dyslexia for the practising optometrist. *Optometry in Practice*, 17, 103-112
19. Wilkins, A.J., Lewis, E., Smith, F., & Rowland, E. (2001). Coloured overlays and their benefit for reading. *Journal of Research in Reading*, 24(1), 41-64.

20. Jeanes, R., Busby, A., Martin, J., Lewis, E., Stevenson, N., Pointon, D. & Wilkins, A. (1997). Prolonged use of coloured overlays for classroom reading. *British Journal of Psychology*, 88, 531– 548.
21. Wilkins, A., Cleave, R., Grayson, N., & Wilson, L. (2009). Typography for children may be inappropriately designed. *Journal of Research in Reading*, 32, 402– 412.
22. Hughes, L.E. & Wilkins, A.J. (2000). Typography in children's reading schemes may be suboptimal: Evidence from measures of reading rate. *Journal of Research in Reading*, 23 (3), 314– 324.
23. Yammouni, R., & Evans, B.J.W. (2020) Is reading rate in digital eyestrain influenced by binocular and accommodative anomalies? *Journal of Optometry*, https://doi.org/10.1016/j.optom.2020.08.006
24. O'Leary, C.I., & Evans, B.J.W. (2006). Double-masked randomised placebo-controlled trial of the effect of prismatic corrections on rate of reading and the relationship with symptoms. *Ophthalmic and Physiological Optics*, 26, 555-565.
25. Lambooij ,M., IJsselsteijn, W.A., Fortuin, M.F., Evans, B.J.W., & Heynderickx, I. (2010). Measuring visual discomfort associated with 3D displays. *J of the SID*, 18: 931– 943.
26. Vander Stappen, C., & Van Reybroeck, C. (2018). Phonological awareness and rapid automatized naming are independent phonological competencies with specific impacts on word reading and spelling: an intervention study. *Front Psychol*, 9, 320.