Vision Research 90 (2013) 43–51

Contents lists available at SciVerse ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Measuring reading performance

Gary S. Rubin *

UCL Institute of Ophthalmology, London, United Kingdom
NIHR Moorfields Biomedical Research Centre, London, United Kingdom

ARTICLE INFO

Article history:
Available online 16 March 2013

Keywords:
Reading
Low vision
Visual impairment
Outcome measures
Clinical trials

ABSTRACT

Despite significant changes in the treatment of common eye conditions like cataract and age-related macular degeneration, reading difficulty remains the most common complaint of patients referred for low vision services. Clinical reading tests have been widely used since Jaeger introduced his test types in 1854. A brief review of the major developments in clinical reading tests is provided, followed by a discussion of some of the main controversies in clinical reading assessment. Data for the Salisbury Eye Evaluation (SEE) study demonstrate that standardised clinical reading tests are highly predictive of reading performance under natural, real world conditions, and that discrepancies between self-reported reading ability and measured reading performance may be indicative of people who are at a pre-clinical stage of disability, but are at risk for progression to clinical disability.

If measured reading performance is to continue to increase in importance as a clinical outcome measure, there must be agreement on what should be measured (e.g. speed or comprehension) and how it should be measured (e.g. reading silently or aloud). Perhaps most important, the methods for assessing reading performance and the algorithms for scoring reading tests need to be optimised so that the reliability and responsiveness of reading tests can be improved.

1. Introduction

In the early 1990s we obtained data from 1000 consecutive patients referred for low vision evaluation at the Johns Hopkins Wilmer Eye Institute low vision service (Unpublished data). An intake questionnaire asked each patient to indicate the primary reason for seeking referral to low vision. The results are shown in Fig 1. The most common reason for referral was difficulty reading, which applied to over 60% of patients. The second most common reason was difficulty driving, applicable to only 5% of patients. Similar results have been published for other populations (see, e.g. Elliott et al., 1997).

Since 1990 there have been significant improvements in the treatment of eye disease – most notably the introduction of anti-VEGF therapy for neovascular (“wet”) AMD. Yet reading difficulty continues to be a primary concern for patients referred for low vision services. In a small but detailed study of patient expectations prior to low vision rehabilitation 14 of 15 patients with AMD reported that reading difficulty was a primary concern (Crossland et al., 2007). Although we are inclined to interpret these findings as an indication of the importance of reading in everyday life, there is another possibility – that patients with reading difficulty are referred to low vision services because low vision rehabilitation is most likely to improve reading performance through the prescription of magnifiers. Other problems such as driving or recognising faces are more difficult to address with current technology and patients with these problems may not be referred.

But in support of the ”reading is important” explanation it is also worth noting that most commonly used questionnaires for assessing the various aspects of vision disability include one or more items on reading difficulty. Popular instruments such as the ADVS (Mangione et al., 1992) VF-14 (Steinberg et al., 1994), NEI-VFQ-25 (Mangione et al., 2001), Massof Activity Inventory (Massof et al., 2005) and many others include an item about difficulty reading newsprint, and entire questionnaires have been developed just to evaluate reading performance such as the Reading Behaviour Inventory (Goodrich et al., 2006). Moreover, measured reading performance is among the best predictors of patient-reported visual ability (McClure et al., 2000) and vision-related quality of life (Hazel et al., 2000).

Reading performance has been used as the primary outcome measure for several clinical trials on the effectiveness of low vision rehabilitation (see Binns et al., 2012) and as a secondary outcome measure for clinical trials of pharmaceutical and surgical treatment of various eye diseases including laser photocoagulation (Macular Photocoagulation Study Group, 1991), submacular surgery (Hawkins et al., 2004), anti VEGF (Tufail et al., 2010) treatments for AMD, and comparison of intraocular lenses following cataract

---

* Address: Department of Visual Neuroscience, UCL Institute of Ophthalmology, London EC1V 9EL, United Kingdom.
E-mail address: g.rubin@ucl.ac.uk

0042-6989/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2013.02.015

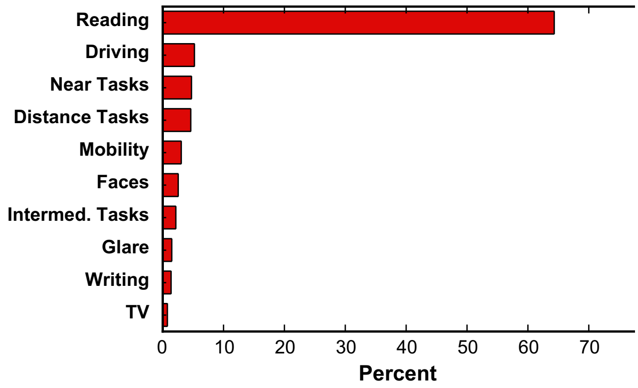


Fig. 1. Chief complaints of 1000 consecutive low vision patients seen at Wilmer Low Vision Service (unpublished data).

extraction (Akutsu et al., 1992). Although reading tests have a long history and extensive literature, there are still several controversial issues about reading ability as a clinical outcome measure. One question is whether standardised tests of reading performance in the lab informs us about reading performance under real-world conditions. A second issue is the relationship between self-reported reading ability and measured reading performance. If the two are in close agreement do we need to measure performance – can't we just ask the patient? And if the two disagree what can we learn from the discrepancy. Finally there are practical questions about how to best measure reading performance. To help put these issues into perspective, it is useful to begin with a brief history of clinical uncireading tests developed for ophthalmic research.

2. A brief history of clinical reading tests

Space does not permit a comprehensive review of reading tests, but the following brief history of these tests highlights some of the key issues about reading assessment that still concern us.

Although clinical reading tests seem to be a relatively recent development, the first known test, developed by Eduard von Jaeger in 1854 (Runge, 2000), actually predated the introduction of Snellen's visual acuity tests in the 1870s (Fig. 2).

The Jaeger test types were based on a graduated series of sentence fragments of decreasing size. In the US, some of the most popular clinical reading charts still specify letter size using the Jaeger J1, J2, etc. notation. The J notation has been criticised for lack of consistency across manufacturers and for the failure to follow a meaningful size progression (Jose & Atcherson, 1977). However the original Jaeger texts followed a strict geometric progression, foretelling the introduction of the Bailey–Lovie Near Reading Card by over 125 years. When the Jaeger charts were first published in the US using local typefaces they lost their original calibration.

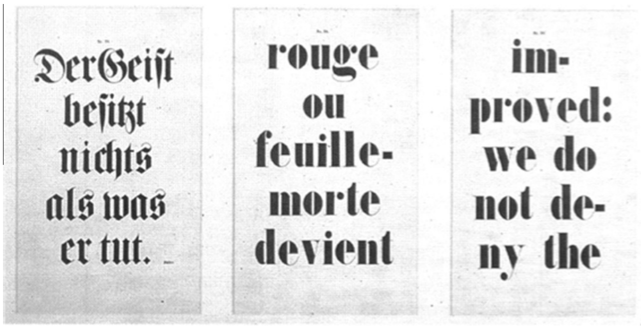


Fig. 2. Original Jaeger test types in German, French and English (from Runge (2000)).

A noteworthy development in clinical reading tests was the Sloan Continuous Text Read Cards, with text size specified in M units (Sloan & Brown, 1963).

Actually, the M unit was promoted and used by Snellen and he tried to convince Jaeger to specify his test types in M units. M notation designates the distance (in metres) at which the object subtends 5 minarc. Therefore 1M print subtends 5 minarc at 1 m. The Sloan reading cards present a short text passage at one size per card (Fig. 3) The amount of text varies with letter size from a few words at 20M to an entire paragraph at 1M. Though popular in low vision clinics, M notation has not been widely adopted elsewhere in clinical ophthalmology.

The next significant advance in reading assessment was the introduction of the Bailey–Lovie Near Reading Card in 1980 (Bailey & Lovie, 1980).

Bailey–Lovie cards present two to six unrelated words per line and the size of the text decreases by a constant percentage from line to line (Fig. 4) Letter size is represented in LogMAR units ($\log_{10}$ of the minimum angle of resolution). Though sometimes criticised because some of the words are quite long (up to 10 letters) and difficult for poor readers, the Bailey–Lovie near cards are still widely used for determining the magnification required to read normal print sizes.

A rather unusual reading test, the Pepper Visual Skills for Reading Test (VSRT) was published in 1986 (Baldasare et al., 1986) by Watson and colleagues at Pennsylvania College of Optometry. The VSRT progresses from well-spaced individual letters, to crowded letters, digrams, trigrams, words and words arranged in a paragraph style (Fig. 5). Unrelated words are used throughout. The test is timed and scored by adding together the number of correct letters, digrams, trigrams, and words read, but the test is said to measure print recognition and navigation skills rather than the amount of magnification required.

Legge and colleagues introduced the MNREAD Test in 1989 (Legge et al., 1989a). Originally a computer-based test, MNREAD was soon converted to printed cards (Fig. 6).

The original MNREAD Test consisted of both sentences and groups of unrelated words rendered in a fixed letter size that subtended 6° at a 20 cm viewing distance. The large print size was designed to measure maximum reading speed rather than reading

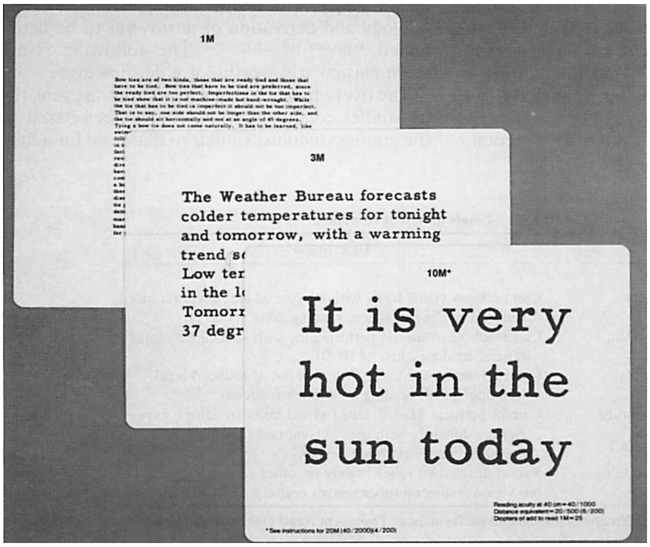


Fig. 3. Louise Sloan's continuous text reading cards with letter size specified in M units (see text).

N LogMAR (at 25 cm)

80 1.6 definition fall

64 flag opening

48 vehicle mark

40 1.3 fellowship tension neat

32 keys earthquake abolish

24 bashful legitimate wood

20 1.0 kennels book negotiated darling percentage cool

16 settlement tidy hundred quip loyalty sanitation

12 army gathers occasional unit generation diploma

10 0.7 liberation nervous moth between jobs vibrations

8 tard tormenting yearned industrial atom forgive

5 0.4

0.0

BAILEY-LOVIE
WORD READING CHART
National Vision Research
Institute of Australia
© Copyright 1979.

Fig. 4. Bailey–Lovie word reading card illustrating logMAR progression of letter sizes.

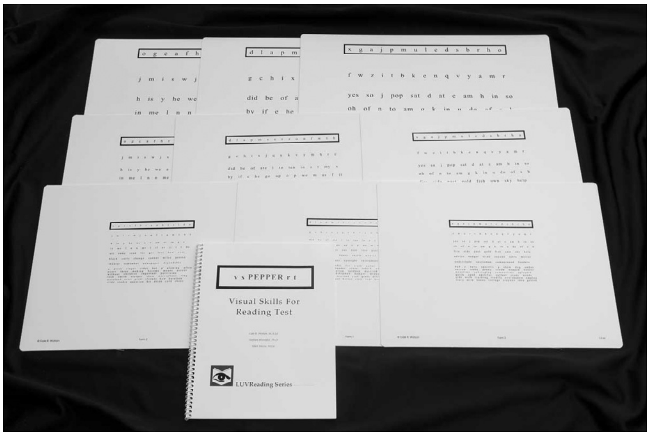
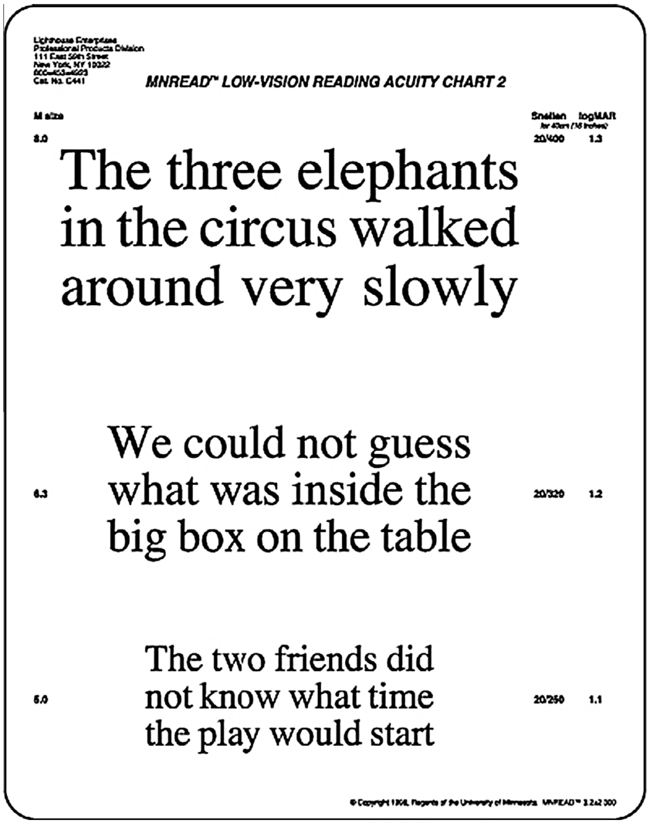


Fig. 5. The Visual Skills for Reading Test (Pepper Test) progresses from single letters to sequences of unrelated words.

acuity. The large print cards were replaced by the MNREAD Acuity Chart, which was designed to measure reading acuity and maximum reading speed (Mansfield et al., 1993; Mansfield, Legge, & Bane, 1996). The MNREAD Acuity Chart consisted of a series of 60-character sentences displayed on two lines. The sentences decrease in size by 0.1 log unit from a maximum of 1.3 logMAR (equivalent to 20/400 or 6/12 when viewed at 40 cm) to –0.5 logMAR (20/6 or 6//2). One advantage of using logMAR scaling of letter size is that the range of print sizes (angular subtense) can be extended by changing the viewing distance.

With the MNREAD Acuity Chart, reading acuity corresponds to the smallest letter size that can be read and maximum reading rate is the number of words read correctly per minute for the sentence with the shortest reading time. A third parameter, critical print size, is the smallest letter size that can be read at the maximum speed and is an indication of the minimum magnification required for best reading, Several variations on the methods of computing maximum reading rate and critical print size have been proposed, (Patel et al., 2011) and these will be discussed below.

Several of the more common reading tests are available in multiple languages. But one test was developed specifically for cross-language comparisons. The International Reading Speed Texts (IReST) are paragraphs of about 170 words (in the English version) that are carefully equated across languages for word frequency and syntactic complexity. Originally published in four European languages, (Hahn et al., 2006) IReST was recently expanded to 17 languages with normative data for normally sighted young adults (Trauzettel-Klosinski, 2012).



Lighthouse Enterprises
Professional Products Division
111 East 59th Street
New York, NY 10022
800-453-4923
Cat. No. C441

MNREAD™ LOW-VISION READING ACUITY CHART 2

M size Snellen logMAR
for 40cm (16 inches)

8.0 20/400 1.3

The three elephants
in the circus walked
around very slowly

6.3 20/320 1.2

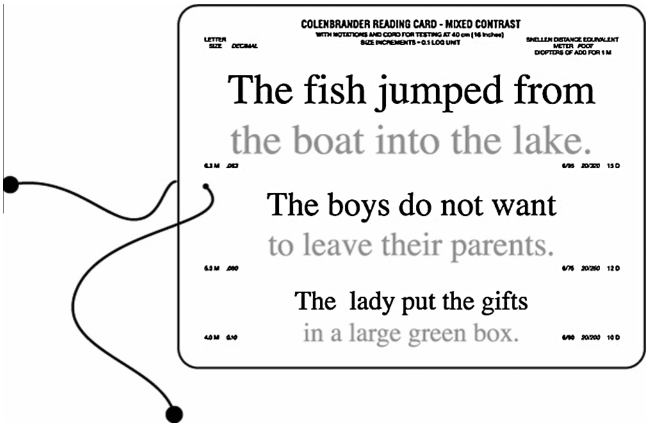
We could not guess
what was inside the
big box on the table

5.0 20/250 1.1

The two friends did
not know what time
the play would start

© Copyright 1994, Regents of the University of Minnesota. MNREAD™ 3.2a2 300

Fig. 6. The MNREAD reading chart consists of standardised sentences displayed in a wide range of letter sizes. The size decreases in a logarithmic fashion with smaller letters on the reverse side of the chart (not shown).



COLENBRANDER READING CARD - MIXED CONTRAST
WITH NOTATIONS AND CORD FOR TESTING AT 40 cm (16 inches)
SIZE INCREMENTS = 0.1 LOG UNIT

LETTER SIZE DECIMAL

SNELLEN DISTANCE EQUIVALENT
METER FOOT
DIOPTERS OF ADD FOR 1 M

The fish jumped from
the boat into the lake.

6.3 M .063 6/96 20/320 15 D

The boys do not want
to leave their parents.

5.0 M .080 6/75 20/250 12 D

The lady put the gifts
in a large green box.

4.0 M 0.10 6/60 20/200 10 D

Fig. 7. The Colenbrander mixed contrast reading card is composed of two-line sentences that follow a logarithmic progression of letter sizes. Lines alternate between high (>90%) and low (10%) contrast.

In addition to the reading tests described above, which use short selections of high-contrast text, there are several special-purpose reading tests that are also worth mentioning. Colenbrander (Dexl et al., 2010) has developed a mixed contrast reading chart with alternating lines of high and low (10%) contrast words (Fig. 7).

The lines decrease in letter size, similar to the Bailey–Lovie card and the test is designed to screen for contrast and reading deficits simultaneously.

A radically different mode of text presentation is used for the RSVP test. The name stands for Rapid Serial Visual Presentation and was first used in 1970 by Forster (1970) to study cognitive processing during reading. With RSVP, single words are presented sequentially at a fixed location on a video display. The sequence is illustrated in Fig. 8. In 1994, we (Rubin & Turano, 1994) introduced RSVP as a means to overcome difficulty generating efficient saccadic eye movements when reading with a non-foveal preferred retinal locus (PRL).

However we observed that people with intact central vision read 2 to 4 times faster with RSVP compared to conventional static presentation while those with central scotomas read only about 40% faster with RSVP (Rubin & Turano, 1994). Eye movement recordings revealed that people with central scotomas still made intra-word saccades when reading with RSVP, presumably because their restricted visual span (Legge et al., 1997) made it difficult to recognise a word with a single fixation. Nevertheless, RSVP continues to be used to isolate visual processing and reduce the influence of eye movements during reading and to control where on the retina text is presented.

Possibly the newest clinical reading test is one designed by Ramulu and colleagues (Ramulu et al., 2013) to evaluate sustained reading. Until recently, all reading tests used relatively brief passages of text – usually no more than 200 words. However, a frequent complaint of readers with low vision is that while they can read a few words or sentences with appropriate magnification, they cannot sustain reading for longer than a few minutes. The new sustained reading test measures reading speed over 30 min of silent reading using 7000-word stories followed by 16–20 comprehension questions. The sustained reading test has been shown to be a valid and reliable measure of sustained reading performance (Ramulu et al., 2013).

The Salzburg Reading Desk (Dexl et al., 2010)s takes a very different approach to measuring reading performance. Instead of presenting text printed on a card or on paper, the SRD displays text on a high-resolution computer monitor (Fig. 9).

One either side of the monitor are IR cameras that capture an image of each pupil and use the distance between pupil centroids to determine viewing distance with much greater accuracy than can be done with a tape measure or knotted length of string. The SRD also has voice detection to accurately measure the beginning and end of a trial. The SRD can display letters, words, and short paragraphs in random order and adjusted to the viewer's preferred letter size or to follow an adaptive staircase technique for efficient measurement of reading acuity and critical print size. However, computer monitors need to be carefully calibrated to ensure that the text is of appropriate luminance and contrast if one wishes to generalise to reading printed text.

3. What do clinical reading tests tell us about reading in the real world?

Clinical reading tests are thoroughly standardised and highly artificial. The content is carefully controlled as are the lighting conditions, viewing distance, letter size and contrast. But when we read at home or while out shopping, all of these factors are allowed to vary. Can we learn anything about real-world reading from standardised laboratory tests?

The Salisbury Eye Evaluation (SEE) Study looked at this question in some detail (West et al., 1997). One hundred participants were selected at random from the original group of 2520 SEE study participants living in Salisbury, MD. All were between the ages of 65 and 85. The participants had been to the SEE clinic to have their vision tested, to answer questionnaires about difficulty with daily activities and to have their reading performance assessed with a computer-based reading test. Short paragraphs (≈100 words) were displayed on the computer monitor for 15 s and the participant read the words aloud. The time to read the text was measured with a stopwatch, the number of words read correctly were counted and reading speed in words/minute was computed. Letter size varied from 0.1° (20/30 or 6/9) to 0.5° (20/120 or 6/36) in equal logarithmic steps.

For the home reading test, participants were asked to read aloud a paragraph selected from a local newspaper. The participant arranged the lighting, chose the viewing distance, and was free to use any vision aids that were customarily used. The results are shown in Fig. 10. The graph plots reading speed at hone as a function of reading speed for the largest print (0.5°) in the clinic.

The correlation is quite high ($r = 0.87$) but the regression line (solid) deviates from the line of equality (dashed). The regression equation.

$$\textit{Home reading rate} = \textit{clinic reading rate} * 0.7 + 24.7.$$

indicates that slower readers do better at home, where they can make full use of whatever adaptations they are accustomed to using. Faster readers do better in the clinic. The reason for this is unclear as we would expect fast readers to be less susceptible to environmental factors such as lighting and show less benefit from the high luminance and high contrast of the clinic test. But the same

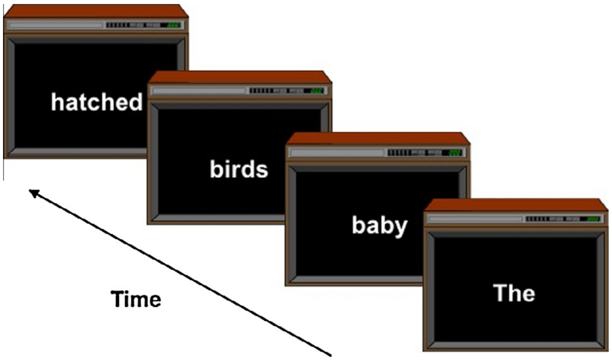


Fig. 8. Demonstration of rapid serial visual presentation. Single words are presented sequentially, centred on a fixed location. RSVP is used to measure reading speed without the need for eye movements.

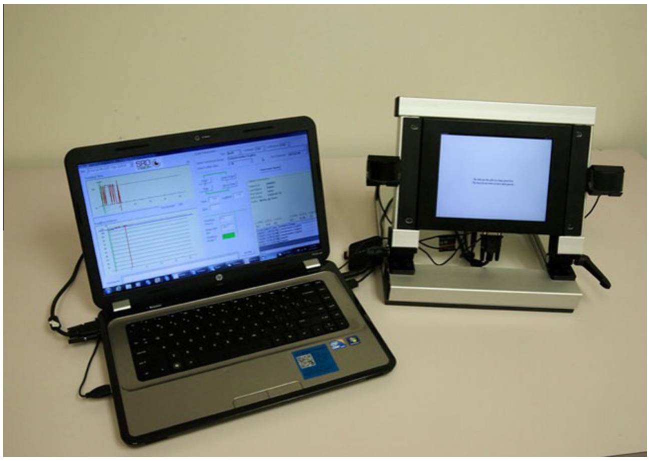


Fig. 9. The Salzburg Reading Desk uses modern computer technology to present text in random order while measuring reading distance with IR cameras and reading speed with voice detection.

effect was observed for other visually demanding tasks such as finding and dialling a phone number.

4. Do we need to measure reading performance? Can't we just ask the patient?

With the current prominence of patient-reported outcome measures it is tempting to conclude that performance-based reading tests are no longer necessary. All we need to do is ask the patient whether he/she has any difficulty reading. However, it has been shown (Guralnik et al., 1989) that performance-based test provide better discrimination in ability level than self report, are earlier predictors of functional decline and disability and are less influence by the participants' sociodemographic, psychosocial, and cognitive characteristics. Also, performance-based tests are independent predictors of morbidity and mortality, even after taking self-report into account.

But how well do patient-reported reading difficulty and measured reading performance agree, and when they disagree does this provide any interesting information about the patient or is it just a reflection of the imprecision of our measurement tools?

Again we can look to the SEE study for some hints (Friedman et al., 1999;). SEE included both patient-reported difficulty reading via the Activities of Daily Vision Scale (ADVS) (Mangione et al., 1992;) and the performance-based reading test described above. The ADVS includes a question about difficulty reading newsprint with response options of "no difficulty", "a little difficulty" "moderate difficulty". "a lot of difficulty," and "can't do" (because of vision problems). Responses to the newsprint question were compared to reading speeds for the text closest in size to newsprint (0.3°). We considered reading speeds greater than 80 words/minute as "functional" reading and reading speeds greater than 160 words/minute as "fluent" (Carver, 1992;). 49.1% of SEE participants reported no difficulty reading and read fluently by our definition, while 3.7% reported at least moderate difficulty reading and read at less than a functional level. In both cases, patient-reported reading difficulty is concordant with measured reading speed. However, 6.4% were slow readers (less than functional reading speed) while reporting no difficulty and 1.5% read fluently while reporting at least moderate reading difficulty. For the majority of participants' self report is in agreement with their measured performance (concordant, unmarked entries in Table 1). But 7.9% show a significant discrepancy between self report and measured reading speed (discordant, single asterisk) and a further 33.8% are mildly discordant (double asterisked entries in Table 1). Some of the discrepancy undoubtedly reflects measurement error, but an analysis of the characteristics of discordant readers (Friedman et al., 1999) suggests a more interesting explanation. When we looked at the vision test results (acuity, contrast sensitivity, glare sensitivity, stereoacuity, and visual fields), all showed a similar pattern of results: visual function for discordant participants was intermediate between results for fast concordant and slow concordant readers. So, for example, distance acuity averaged −0.04 logMAR (±04 S.E) for fast concordant readers (read fluently and report no difficulty), 0.15 logMAR (±0.01) for slow discordant readers (slow readers who report no difficulty) and 0.40 logMAR (±).02 S.E.) for slow concordant reader (read slowly and report difficulty). Furthermore, 80% of discordant readers showed concordance between measured performance and self report when reading text of a larger print size.

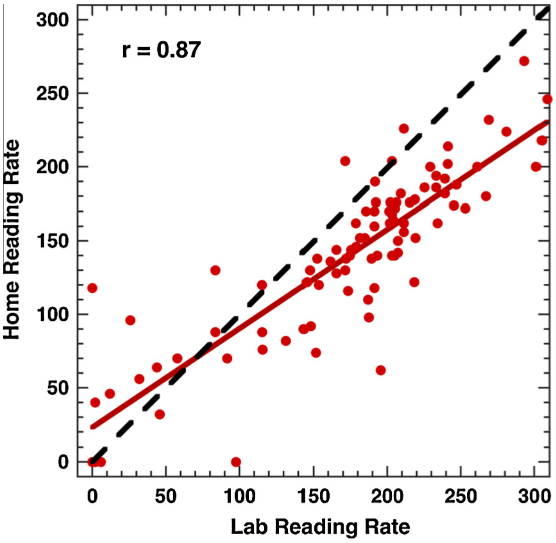


Fig. 10. Comparison of reading rate under standardised laboratory conditions to reading rate under natural conditions at home. Solid line is least squares regression line. Dashed line indicates equality between lab and home.

Taken together, these results suggest that a discrepancy between performance-based tests and self report may be indicative of patients who are at a transition between visual ability and disability where visual function has begun to decline but the person is able to maintain (or at least thinks they can maintain) good performance, possibly through modification of the task. In the geriatrics literature this is referred to as "preclinical" disability and is an important predictor of future disability if left unattended (Fried et al., 1991).

The association of visual acuity with concordance/discordance described above does not mean that a simple test of letter acuity will substitute for measuring reading performance. In a study of 40 patients with AMD, visual acuity was not correlated with reading speed, even for text that was magnified to greater than the critical print size ($r = 0.26$, $p > 0.1$ (Rubin & Feely, 2009)).

5. How should we measure reading performance?

If we accept that clinical reading tests are informative about everyday reading outside the clinic, and that the measurement of reading performance provides additional information that is not captured by self-report alone, then we must ask how should that performance be measured? As the review above makes clear, there are many different types of reading tests. It is natural to ask which test is "the best." However, the optimal test will depend on how it is to be used. If an investigator wants to know whether a pharmaceutical treatment retains or restores vision, as measured by the ability to read small print, then a test with multiple print sizes held at a fixed distance (such as MNREAD) may be most suitable. But if the investigator needs to evaluate how well a patient reads ordinary text with available low vision aids then a test with longer passages of fixed print size (such as IReST) viewed from a distance that is appropriate for the low vision aid may be more appropriate. Nevertheless, there are certain well-accepted standards for comparing

Table 1
Comparison of Self-reported reading difficulty with measured reading speed.

Measured reading speed	Self-reported difficulty reading newsprint (%)		
	Moderate	A little	None
Slow (<80 words/min)	3.7	*3.5*	*6.4***
Functional ($80 \leqslant$ words/min < 160)	*2.1**	5.4	21.3*
Fluent ($\geqslant$160 words/min)	*1.5***	*6.9**	49.1

Unmarked values are concordant, in italics with double asterisks are strongly discordant, and in italics with single asterisk are mildly discordant.

and selecting among tests. These are based on demonstration of the test's validity (does the test measure what it is intended to measure?), reliability (are the measurements consistent and repeatable?) and responsiveness (is the test able to measure change?). Tests used for diagnostic purposes also need to be evaluated for sensitivity and specificity, but since we are not proposing that reading tests be used to aid diagnosis, sensitivity and specificity are of less importance.

None of the reading tests has been thoroughly evaluated for validity, reliability, and responsiveness in visually impaired readers. In most cases, the evaluation has been restricted to test–retest variability and often limited to readers with normal vision. Few studies have made direct comparisons between tests and comparing across studies is difficult when the testing conditions and subject characteristics differ. Clearly, more data are needed to determine the psychometric properties of available reading tests.

Despite 150 years of development and refinement of clinical reading tests, there are still several points of disagreement. The first is what should be measured? In developing the scoring algorithm for the MNREAD Test, Legge and colleagues (Mansfield, Legge, & Bane, 1996) defined three parameters: reading acuity (the smallest print that can be read, however slowly), maximum reading rate (the fastest reading rate regardless of print size) and critical print size (the smallest letter size that allows reading at the maximum rate). There is little controversy about reading acuity. Following Bailey's recommendation for scoring letter acuity charts, reading acuity is scored by counting the number of words read correctly, until the participant no longer identify the text, and the count is converted to a LogMAR value that takes viewing distance into account. Maximum reading rate and critical print size are not so simple. There at least four methods for calculating maximum reading rate and four for critical print size. The various methods are described and compared in a recent paper (Patel et al., 2011) and there is not space here for a thorough discussion of the pros and cons of each method. Briefly, most of the definitions rely on an underlying model for the shape of the reading rate vs. letter size function. This function is thought to rise rapidly from 0 words/minute at the reading acuity until it reaches a plateau at the maximum reading rate. The critical print size is at the "knee" between the rising part of the function and the plateau. Real data show that patients who have very poor vision may fail to reach a plateau and even for patients with good vision, it is sometimes difficult to discern which points belong to the plateau., The uncertainty results, in part, from imprecision in the measurement of reading speed when using short, 60 character sentences. The reaction time of the experimenter when using a stop watch to time each sentence, pauses, false starts, time taken to self-correct reading errors, and other "glitches" by the reader, all lessen the precision and repeatability of reading speed measurements. A study of the test–retest variability of the MNREAD Test with a group of AMD patients participating in a clinical trial of anti-VEGF therapy (Patel et al., 2011) reported coefficients of repeatability of 0.30 logMAR for reading acuity, about 0.55 logMAR for critical print size, and more than 60 words/minute for maximum reading rate. The exact values depended on the definition of maximum reading rate and critical print size used. Another study conducted in a laboratory setting with highly trained researchers and less fatigued patients produced much better coefficients of repeatability (0.1 logMAR, 0.2 logMAR and 10 words/minute for reading acuity, critical print size and maximum reading rate; (Subramanian & Pardhan, 2009)). One approach to this problem has been to apply a statistical model to the analysis, such as the nonlinear mixed effects model of Cheung et al. (2008). NLME has been applied successfully to data from AMD patients, but not to other types of patients. Moreover, there is no simple, practical means of processing MNREAD data with NLME for those who are unfamiliar with R programming. Therefore, most clinical studies that use MNREAD follow either the manufacturer's instructions or one of the published variants.

Another option, for those interested only in reading speed, is to use longer passages of text that are less susceptible to "glitches" in timing. One such test is the International Reading Speed Texts (Trauzettel-Klosinski, Dietz, & Group, 2012), mentioned above, which consists of ten 170-word paragraphs. With ten paragraphs the IReST can be used in clinical trials with several follow up exams without repeating the text. So far repeatability data have only been published for young readers with normal vision.

So far the discussion has centred on factors related to letter size and reading speed. There are other factors, which may be important, such as comprehension and endurance. Comprehension is of obvious importance, but it is seldom measured in the context of clinical vision research. Watson argues that readers with low vision need to relearn cognitive as well as visual processing skills, and that most reading tests ignore this aspect of vision rehabilitation, to the detriment of low vision patient (Watson, 1992). However, a study by (Legge et al. (1989b)) showed that most readers with low vision maintain normal levels of comprehension at reading rates up to 85% of their maximum reading rate, and a study of reading with RSVP (Rubin & Turano, 1992) demonstrated that readers who could accurately repeat sentences presented with RSVP, comprehended what they had read even if the text was presented at much faster rates than they were able to read conventional static text. These studies suggest it is unlikely that readers would pass the speed criteria for fluent reading, but fail to comprehend what they had read. If this is true, then it is questionable whether a test of reading comprehension adds important information to the clinical assessment of reading performance.

Reading endurance is a different matter. As mentioned above, Ramulu and colleagues (2013) have recently developed and validated a test of reading endurance using 7000-word passages followed by 16–20 comprehension questions that can only be answered by reading the passage and are not based on general knowledge. The new silent reading test is a more sensitive indicator of reading difficulty than the standard reading aloud in patients with ocular conditions as diverse as glaucoma and ptosis. However, the test takes up to 30 min, and it is likely to be reserved for reading studies where endurance and fatigue are of particular interest and not as a routine clinical outcome measure.

A second broad question is how should reading speed be measured? Should we use continuous text or unrelated words, read silently or aloud? Semantic context plays an important role for experienced fluent readers. One argument is that reading performance for meaningful text involves complex non-visual factors that are minimised when reading random words. There has been some controversy whether readers with low vision show the same benefit from sentence context. The argument is that low-vision readers who must struggle to decode the visual information may not have sufficient cognitive reserve to take full advantage of semantic context. In two studies that looked specifically at this issue, readers with central field loss (Fine & Peli, 1996) and normally-sighted observers (Fine et al., 1999) forced to use peripheral vision to read showed the same benefit of semantic context when reading meaningful text rather than random word lists. However a study by Sass and colleagues (Sass, Legge, & Lee, 2006) found that normally-sighted readers were better able to use context than readers with low vision. In any event, reading studies using visually degraded text show that the effects of the degradation are amplified when the words are presented within a semantic context (Becker & Killion, 1977).

The controversy over semantic context highlights the fact that reading performance depends on cognitive, linguistic, and motivational factors; not just vision. Although we tend to ignore these

Table 2
Advantages and disadvantages of MNREAD Acuity Test.

MNREAD Acuity Test	
Advantages	Disadvantages
It allows the investigator to extract the three important parameters: reading acuity, maximum reading rate, and critical print size Letter sizes follow a logarithmic progression Sentences are standardised for reading level and length Available in a range of languages Good test–retest variability when both tests conducted on the same day by one experienced examiner	Only 2 charts are available per language so sentences will need to be repeated if used for longitudinal studies Short sentences may be difficult to accurately time and are susceptible to reading "glitches" such as false starts, time taken to self-correct reading errors, both of which may increase test–retest variability Somewhat awkward to hold – the examiner needs three hands for a stopwatch, score sheet, and to maintain a standard viewing distance Requires calibrated external lighting which may be difficult to reproduce outside the lab Sophisticated scoring software is not readily available to users unfamiliar with R programming Poorer test–retest variability when tests conducted on separate days at the end of lengthy clinical trial visits by multiple examiners

Table 3
Advantages and disadvantages of IReST Test.

IReST Test	
Advantages	Disadvantages
Available in many languages (17 at present)	Available only in one size – Times Roman 12 pt. for languages using the Roman alphabet
Careful standardisation of linguistic complexity across languages makes it possible to do multinational comparisons	Multiple text passages per card might confuse some readers, especially those using magnifications devices
Ten texts make it possible to do longitudinal studies without repeating passages	
Texts sufficiently long (170 words) to minimise the effect of reading "glitches" which should improve test–retest variability	
Low within-subject variability supports good reliability, but only tested in young normally-sighted readers	No data on test–retest variability patients or elderly readers

other factors, they can have dramatic and complex effects on measured reading performance. To minimise the influence of cognitive reading ability, it is important to select text at the appropriate reading level. Carver contends that cognitive reading ability exerts little influence on reading speed if the participant's reading level is at least three grades above the grade level of the text (Carver, 1992). Most reading tests use text at Grade 6 or below (US) which should provide the necessary margin. Duchnicky and Kolers claim that reading speed is less sensitive to cognitive factors and more sensitive to vision than reading comprehension, providing another argument in favour of measuring speed (Duchnicky & Kolers, 1983).

Should reading performance be assessed by reading aloud or reading silently? Practically, it is much more difficult to evaluate reading speed when reading silently. Without resorting to comprehension tests it is difficult to insure that silently read text is accurately read; not just skimmed. Although silent reading is generally faster than reading aloud, both forms of reading are similarly affected by changing letter size (Chung, Mansfield, & Legge, 1998) and both are predicted by the same clinical tests (Lovie-Kitchin, Bowers, & Woods, 2000).

In addition to these fundamental questions about the best way to evaluate reading performance, there are several subsidiary questions about text layout and presentation that may influence the choice of a reading test.

Font. It has long been argued whether the font used for the reading test makes a difference. The evidence shows that font per se makes little difference to reading speed. The apparent advantage of one font over another can often be traced to differences in stoke width, inter-letter spacing, or the designation of letter size whereby two fonts that are nominally the same size (e.g. both 12 pt) differ in actual size and the amount space they occupy (Rubin et al., 2006).

Spacing between letters. Reading performance is strongly affected by crowding between letters (Pelli et al., 2007). Because crowding effects increase with distance from the fovea, low-vision readers with central scotomas are expected to be especially sensitive to crowding effects and it has been hypothesised that increasing the inter-letter spacing beyond the normal range would improve reading performance in these patients. However, experimental studies have shown that "normal" spacing is optimal and there is little advantage to increased spacing (Chung, 2002).

Word length. Word length is related to text complexity (reading level). However, most reading tests aim for a reading level at or below grade 6 (in the US) and if the experimenter is concerned that the text may vary in reading level, this can be factored out of the reading assessment by converting reading speed to characters/second instead of words/minute (Carver, 1992). This also helps equate reading speeds across languages (Hahn et al., 2006).

6. Conclusion

Improving reading ability is a high priority for patients threatened with the loss of vision. Reading speed is a strong predictor of visual ability and vision-related quality of life. From this, we would expect reading performance to be one of the more important outcome measures for judging the effectiveness of therapeutic interventions and vision rehabilitation. But that is not yet the case. In the century and a half since the introduction of Jaeger's first clinical reading test, there have been dozens, if not hundreds, of different reading tests. But there is not yet a consensus on the best way to evaluate reading performance, as there is for visual acuity (logMAR letter charts) and contrast sensitivity (variable contrast letter charts). But most tests have settled on a set of common features: (1) reading speed is the key outcome variable, with tests of comprehension or reading endurance reserved for specific research questions, (2) reading aloud is preferred for ease of scoring (3) reading speed is measured for meaningful text even though this may allow greater influence of cognitive factors.

When we are interested in measuring reading speed across a range of letter sizes, the MNREAD Acuity Test is a popular choice. Its advantages and disadvantages are listed in Table 2.

For measuring reading speed for a standard print size, the IReST has several advantages, but some disadvantages, listed in Table 3.

Despite the many differences between highly standardised clinical reading tests and normal, everyday reading, performance on the clinical tests is highly predictive of everyday reading. Most visual function questionnaires include a patient-reported assessment of reading difficulty and while the self-reported ability usually agrees with measured reading performance, there may be differences, particularly when the patient reads slowly but reports no difficulty, which could be indicative of pre-clinical disability Table 1.

A central concern for those thinking about using some form of reading assessment in their next clinical trial is the questionable reliability of current reading tests. Outcome measures with poor reliability inflate sample sizes required to detect treatment effects. More research is needed to optimise reliability of clinical reading tests. With the advent of new technology for reading – e-book readers, tablets and notebook computers with improved resolution – there is likely to be a change in technology for reading assessment that may help address this issue.

Acknowledgments

The writing of this manuscript was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

Akutsu, H., Legge, G. E., Showalter, M., Lindstrom, R. L., Zabel, R. W., & Kirby, V. M. (1992). Contrast sensitivity and reading through multifocal intraocular lenses. *Archives of Ophthalmology, 110*, 1076–1080.

Bailey, I. L., & Lovie, J. E. (1980). The design and use of a new near-vision chart. *American Journal of Optometry and Physiological Optics, 57*, 378–387.

Baldasare, J., Watson, G. R., Whittaker, S. G., & Miller-Shaffer, H. (1986). The development and evaluation of a reading test for low vision individuals with macular loss. *Journal of Visual Impairment & Blindness*, 785–789.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception, 3*, 389–401.

Binns, A. M., Bunce, C., Dickinson, C., Harper, R., Tudor-Edwards, R., Woodhouse, M., et al. (2012). How effective is low vision service provision? *A systematic review. Survey of Ophthalmology, 57*(1), 34–65.

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, 84–95.

Cheung, S. H., Kallie, C. S., Legge, G. E., & Cheong, A. M. (2008). Nonlinear mixed-effects modeling of MNREAD data. *Investigative Ophthalmology & Visual Science, 49*(2), 828–835.

Chung, S. T. L. (2002). The effect of letter spacing on reading speed in central and peripheral vision. *IOVS, 43*(4), 1270–1276.

Chung, S., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research, 38*, 2949–2962.

Crossland, M. D., Gould, E. S., Helman, C. G., Feely, M. P., & Rubin, G. S. (2007). Expectations and perceived benefits of a hospital-based low vision clinic: Results of an exploratory, qualitative research study. *Visual Impairment Research, 9*(2–3), 59–66.

Dexl, A. K., Schlogel, H., Wolfbauer, M., & Grabner, G. (2010). Device for improving quantification of reading acuity and reading speed. *Journal of Refractive Surgery, 26*(9), 682–688.

Duchnicky, R. L., & Kolers, P. A. (1983). Readability of text scrolled on visual display terminals as a function of window size. *Human Factors, 25*, 683–692.

Elliott, D. B., Trukolo-Ilic, M., Strong, J. G., Pace, R., Plotkin, A., & Bevers, P. (1997). Demographic characteristics of the vision-disabled elderly. *Investigative Ophthalmology & Visual Science, 38*(12), 2566–2575.

Fine, E. M., Hazel, C. A., Petre, K. L., & Rubin, G. S. (1999). Are the benefits of sentence context different in central and peripheral vision? *Optometry and Vision Science, 76*(11), 764–769.

Fine, E. M., & Peli, E. (1996). The role of context in reading with central field loss. *Optometry and Vision Science, 73*, 533–539.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Attention, Perception, & Psychophysics, 8*, 215–221.

Fried, L. P., Herdman, S. J., Kuhn, K. E., Rubin, G., & Turano, K. (1991). Preclinical disability: Hypotheses about the bottom of the iceberg. *Journal of Aging and Health, 3*, 285–300.

Friedman, S. M., Munoz, B., Rubin, G. S., West, S. K., Bandeen-Roche, K., & Fried, L. P. (1999). Characteristics of discrepancies between self-reported visual function and measured reading speed. Salisbury Eye Evaluation Project Team. *Investigative Ophthalmology & Visual Science, 40*(5), 858–864.

Goodrich, G. L., Kirby, J., Wood, J., & Peters, L. (2006). The reading behavior inventory: An outcome assessment tool. *Journal of Visual Impairment & Blindness, 100*(3), 164–168.

Guralnik, J. M., Branch, L. G., Cummings, S. R., & Curb, J. D. (1989). Physical performance measures in aging research. *Journal of Gerontology, 44*, M141–M146.

Hahn, G. A., Penka, D., Gehrlich, C., Messias, A., Weismann, M., Hyvarinen, L., et al. (2006). New standardised texts for assessing reading performance in four European languages. *British Journal of Ophthalmology, 90*(4), 480–484.

Hawkins, B. S., Miskala, P. H., Bass, E. B., Bressler, N. M., Childs, A. L., Mangione, C. M., et al. (2004). Surgical removal vs observation for subfoveal choroidal neovascularization, either associated with the ocular histoplasmosis syndrome or idiopathic: II. Quality-of-life findings from a randomized clinical trial: SST Group H Trial: SST Report No. 10. *Archives of Ophthalmology, 122*(11), 1616–1628.

Hazel, C. A., Latham, P. K., Armstrong, R. A., Benson, M. T., & Frost, N. A. (2000). Visual function and subjective quality of life compared in subjects with acquired macular disease. *Investigative Ophthalmology and Vision Science, 41*, 1309–1315.

Jose, R. T., & Atcherson, R. M. (1977). Type-size variability for near-point acuity tests. *American Journal of Optometry and Physiological Optics, 54*(9), 634–638.

Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading-XVI. The visual span in normal and low vision. *Vision Research, 37*, 1999–2010.

Legge, G. E., Ross, J. A., Luebker, A., & LaMay, J. M. (1989a). Psychophysics of reading. VIII. The Minnesota low-vision reading test. *Optometry and Vision Science, 66*, 843–853.

Legge, G. E., Ross, J. A., Maxwell, K. T., & Luebker, A. (1989b). Psychophysics of reading – VII: Comprehension in normal and low vision. *Clinical Vision Sciences, 4*, 51–60.

Lovie-Kitchin, J. E., Bowers, A. R., & Woods, R. L. (2000). Oral and silent reading performance with macular degeneration. *Ophthalmic & physiological optics: The Journal of the British College of Ophthalmic Opticians, 20*(5), 360–370.

Macular Photocoagulation Study Group (1991). Laser photocoagulation of subfoveal neovascular lesions in age-related macular degeneration: Results of a randomized clinical trial. *Archives of Ophthalmology, 109*, 1220–1231.

Mangione, C. M., Lee, P. P., Gutierrez, P. R., Spritzer, K., Berry, S., & Hays, R. D. (2001). Development of the 25-item national eye institute visual function questionnaire. *Archives of Ophthalmology, 119*(7), 1050–1058.

Mangione, C. M., Phillips, R. S., Seddon, J. M., Lawrence, M. G., Cook, E. F., Dailey, R., et al. (1992). Development of the 'Activities of Daily Vision Scale': A measure of visual functional status. *Medical Care, 30*, 1111–1126.

Mansfield, J. S., Ahn, S. J., Legge, G., & Luebker, A. (1993). A new reading acuity chart for normal and low vision. In *Noninvasive assessment of the visual system: summaries. Technical digest series* (pp. 232–235). Optical Society of America.

Mansfield, J. S., Legge, G. E., & Bane, M. C. (1996). Psychophysics of reading. XV: Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science, 37*(8), 1492–1501.

Massof, R. W., Hsu, C. T., Baker, F. H., Barnett, G. D., Park, W. L., Deremeik, J. T., et al. (2005). Visual disability variables. I: The importance and difficulty of activity goals for a sample of low-vision patients. *Archives of Physical Medicine and Rehabilitation, 86*(5), 946–953.

McClure, M. E., Hart, P. M., Jackson, A. J., Stevenson, M. R., & Chakravarthy, U. (2000). Macular degeneration: Do conventional measurements of impaired visual function equate with visual disability? *British Journal of Ophthalmology, 84*, 244–250.

Patel, P. J., Chen, F. K., Da Cruz, L., Rubin, G. S., & Tufail, A. (2011). Test–retest variability of reading performance metrics using MNREAD in patients with age-related macular degeneration. *Investigative Ophthalmology & Visual Science, 52*(6), 3854–3859.

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision, 7*(2), 20, 21–36.

Ramulu, P. Y., Swenor, B. K., Jefferys, J. L., & Rubin, G. S. (2012). Description and validation of a test to evaluate sustained silent reading. *Investigative Ophthalmology & Visual Science, 54*(1), 673–680.

Rubin, G. S., & Feely, M. (2009). The role of eye movements during reading in patients with age-related macular degeneration (AMD). *Neuro-Ophthalmology, 33*(3), 120–126.

Rubin, G. S., Feely, M., Perera, S., Ekstrom, K., & Williamson, E. (2006). The effect of font and line width on reading speed in people with mild to moderate vision loss. *Ophthalmic and Physiological Optics, 26*(6), 545–554.

Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye movements. *Vision Research, 32*, 895–902.

Rubin, G. S., & Turano, K. (1994). Low vision reading with sequential word presentation. *Vision Research, 34*, 1723–1733.

Runge, P. E. (2000). Eduard Jaeger's Test-Types (Schrift-Scalen) and the historical development of vision tests. *Transactions of the American Ophthalmological Society, 98*, 375–438.

Sass, S. M., Legge, G. E., & Lee, H. W. (2006). Low-vision reading speed: Influences of linguistic inference and aging. *Optometry and Vision Science, 83*(3), 166–177.

Sloan, L. L., & Brown, D. J. (1963). Reading cards for selection of optical aids for the partially sighted. *American Journal of Ophthalmology, 55*, 1187–1199.

Steinberg, E. P., Tielsch, J. M., Schein, O. D., Javitt, J. C., Sharkey, P., Cassard, S. D., et al. (1994). The VF-14, an index of functional impairment in patients with cataract. *Archives of Ophthalmology, 112*, 630–638.

Subramanian, A., & Pardhan, S. (2009). Repeatability of reading ability indices in subjects with impaired vision. *Investigative Ophthalmology & Visual Science, 50*(8), 3643–3647.

Trauzettel-Klosinski, S., Dietz, K., & Group, I. R. S. (2012). Standardized assessment of reading performance: The new international reading speed texts IReST. *Investigative Ophthalmology & Visual Science, 53*(9), 5452–5461.

Tufail, A., Patel, P. J., Egan, C., Hykin, P., da Cruz, L., Gregor, Z., et al. (2010). Bevacizumab for neovascular age related macular degeneration (ABC Trial): Multicentre randomised double masked study. *BMJ, 340*, c2459.

Watson, G. R. (1992). The efficacy of comprehension training and reading practice for print readers with macular loss. *Journal of Visual Impairment and Blindness, 86*(1), 37–43.

West, S. K., Rubin, G. S., Munoz, B., Abraham, D., & Fried, L. P. (1997). Assessing functional status: correlation between performance on tasks conducted in a clinic setting and performance on the same task conducted at home. The Salisbury Eye Evaluation Project Team. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 52*(4), M209–M217.