Comparing Reading Speeds for Reading Standardized Single Sentences and Paragraphs in Patients with Maculopathy

Constanze Kortuem^a Tobias Marx^b Elke Karin Altpeter^b
Susanne Trauzettel-Klosinski^c Stephan Kuester-Gruber^c

^aCenter for Ophthalmology, Department for Strabology, Lids and Orbit, University of Tuebingen, Tuebingen, Germany; ^bCenter for Ophthalmology, Low Vision Clinic, University of Tuebingen, Tuebingen, Germany; ^cCenter for Ophthalmology, Vision Rehabilitation Research Unit, University of Tuebingen, Tuebingen, Germany

Keywords

Reading speed · Variability · Repeatability · Age-related macular degeneration · Low vision · Maculopathy · Sentences · Paragraphs

Abstract

Introduction: The purpose of this study was to compare reading speeds (RS) of either paragraphs of text or single sentences in patients with maculopathy by investigating the repeatability of RS to decide which text type to use in studies on patients with maculopathy. **Methods:** RS was measured for standardized texts in 25 participants with a central scotoma due to maculopathy (mean age 77.8 years $\pm$ 9.9 SD, mean binocular visual acuity 0.65 logMAR ($\pm$ 0.85 SD), median magnification requirement 3-fold). Reading 3 single sentences taken from the Radner reading charts (sample #1, #2, and #3) of 14 words and reading 3 paragraphs of International Reading Speed Texts (IReST, sample # 3, #6, and #10) with a mean of 132 ($\pm$ 3.2 SD) words, each in German, were compared. The 6 texts were read aloud in random order from a closed-circuit TV system, with size adjusted according to the individual magnification requirement. Reading time was measured by stopwatch, and speed was calculated in correctly read words per minute (wpm). Differences in RS depending on text length (single sentence vs. paragraph) and

text sample were calculated by the Bland-Altman analysis. **Results:** The mean RS showed no significant difference between 2 charts of the same kind (sentences: 93 wpm $\pm$ 37 SD; paragraphs: 95 wpm $\pm$ 38 SD). RS differences between 2 charts were lower in paragraphs than in single sentences. Highest correlations of RS between all 6 texts existed between the 3 IReST text samples (r = 0.98, 0.98, and 0.98) compared with the 3 Radner sentences (r = 0.89, 0.81, and 0.90). The inter-chart reliability (coefficient of repeatability) was smaller for the paragraphs (12.9 wpm) than for the single sentences (36.4 wpm). **Conclusion:** In patients with maculopathy, single sentences are well suited for single measurement of RS. For repeated measurements (e.g., monitoring the course of a reading disorder or assessing effects of interventions), paragraphs are preferable because of their lower variability of RS between the paragraphs.

© 2020 The Author(s).
Published by S. Karger AG, Basel

Introduction

To test reading speed (RS) in a clinical and research setting, standardized reading texts are necessary. Such texts are available as single sentences to measure reading acuity and critical print size, and also RS, such as MNread [1], Colenbrander [2], SKread [3, 4], and the Radner Reading

Charts [5, 6]. On the other hand, there are also longer paragraphs to measure RS, such as the International Reading Speed Texts (IReST) [7, 8] and the Radner paragraphs [9]. Reading a paragraph resembles real-life reading more than reading a single sentence, but the measurement is more time-consuming. The IReST were the first standardized paragraphs for measuring RS [7]. They were developed in order to take advantage of the fact that several sentences in a paragraph in fact average the RS of the constituent sentences, which would result in reduced variability. In addition, they are also supposed to approximate the everyday life reading situation, like reading a passage in a newspaper. Both expectations were confirmed in a previous study on normally sighted participants [10]. No significant difference in mean RS between the 2 types of text length was found, but RS varied less for different samples of IReST paragraphs [10]. Brussee et al. [11] compared the Radner charts and the IReST charts in Dutch and also found a lower variability of reading longer texts than single sentences in normally sighted participants.

The aim of this study was to investigate whether the smaller variability between text samples in the IReST paragraphs compared with single sentences observed in normally sighted older adults [10] can also be found in older patients with maculopathy who use magnifying visual aids. The IReST were specifically developed for measuring RS rather than reading acuity and critical print size to serve as a supplement to previously existing reading tests. Therefore, the IReST charts are provided in only 1 text size (1.25 M), the common newspaper print, which is the standard requirement for normally sighted subjects and patients with low vision. Therefore, low vision patients should read them with their magnifying visual aids in the clinical setting, which allows the medical staff to choose the best visual aid and to judge the effect during the phase of its individual adjustment. This is important because reading with their magnifying visual aids corresponds to the patients' everyday situation.

Additionally, the IReST were developed as a set of 10 equivalent paragraphs that provide better preconditions for repeated measurements, that is, 10 variants can be used on a patient before having to repeat previously seen and possibly remembered text. All 10 paragraphs have the same length, difficulty, and linguistic complexity [12]. The same rules (same content, length, difficulty, and linguistic complexity) were applied to the additional 19 languages to make the texts comparable among languages for international studies [8, 13]. Therefore, the findings based on the German texts used here are not limited to this language alone.

Recently, Radner et al. [9] presented another type of paragraphs for RS analysis. The texts are standardized regarding sentence construction as well as number, length, and position of words. Words with the same number of syllables appear in identical positions in the text. This means that the texts are equal regarding construction. However, these sentences sometimes result in unusual word sequences. RS depends not only on the sentence construction in the sense of identical word length and position, but also on the degree of difficulty, which is influenced by word frequency [14], linguistic complexity [12], and the degree of interest that the text can evoke in the reader.

In everyday life, we read almost always silently. Studies have shown that RS in this reading mode is faster compared with reading out loud [15, 16]. However, testing with silent reading requires questions to verify comprehension. Even though standardized texts read out loud may be artificial [17], this reading mode offers the advantage of automatically revealing mistakes, as well as the exact beginning and end of the reading process. Furthermore, using a long reading time (30 min), as in the study by Ramulu et al. [16] is helpful for measuring fatigue, but it is not practical in a routine clinical situation as the time available for patient examinations is limited.

RS is faster in younger than in older adults, and advanced age has been found to be a major factor in determining RS [18]. In a previous study on normally sighted participants using IReST [7], the young participants (18–35 years) were 19% faster than the older adults (60–85 years). In a subsequent study [8], mean RS in young participants (18–35 years) was 179 words per minute (wpm), with a normal lower reference value of 142 wpm (2.5% quantile of RS). This is why for better comparison normally sighted older adults (mean age 64.5 years; mean RS: 167 wpm) were examined in a previous study [10] with the same methods described here.

Compared with subjects with normal vision, RS is markedly lower in patients with central visual field defects [18–21]. In a previous study, a mean RS of participants with age-related macular degeneration (AMD) of 74 wpm was found [22]. Patients with a central scotoma have to overcome the fact that they are forced to read with an eccentric retinal locus, also called preferred retinal locus (PRL). It stands to reason that the size of the scotoma and, thus, the location of the PRL are important factors that determine their RS [19, 23]. An explanation for the influence of central visual field defect characteristics has been hypothesized on the basis of the shrinking perceptual span [24, 25]. It has also been found that the type of

AMD (i.e., wet vs. dry) is another factor to be considered [21]. However, in the present study, this differentiation was not an issue, because all participants had a stable maculopathy in a dry stage at the time of the study. As any eccentric PRL has insufficient resolution for reading normal print size, the text has to be magnified (for an overview, see Trauzettel-Klosinski [26, 27]).

The measurement of visual acuity with single optotype does not reflect the ability to read, because it neglects the necessity to plan and execute eye movements to achieve reading fluency. Based on these facts, one can speculate that the considerable effort that has to be made by a patient with a central scotoma may cause considerable strain. This could be exhausting, as it can affect reading performance more while reading longer paragraphs than while reading a single sentence, which can be exacerbated by advanced age.

It was hypothesized that participants with AMD will read text in a length of several sentences at a slower speed than single sentences. If this is true, one could conclude that the effect is due to fatigue, which is more likely to occur during longer trials. The sentences have only a few words per line, whereas the paragraphs have longer lines. Shorter lines require shorter return sweeps, which can be especially helpful to individuals with AMD if the beginning of the next line does not fall in the scotoma. Therefore, the text used for reading paragraphs with longer lines could be more demanding than the text used for single sentences because they require larger return sweeps.

To the best of our knowledge, this is the first study that directly compared RS between standardized reading texts of paragraphs versus single sentences in patients with maculopathy. For comparison, we chose the Radner single sentence charts in German (edition 1998), because they are standardized and well established and because they were applied in a previous study on normally sighted participants [10]. Part of these data were published in the doctoral thesis by T.M. in German [28].

Materials and Methods

Subjects

Approval for this study was obtained from the Ethics Committee of the Medical Faculty, University of Tuebingen, Germany, on April 18, 2012. The study was registered in the German Clinical Trials Register (DRKS-ID 00021615). All procedures complied with the tenets of the Declaration of Helsinki. Twenty-five participants were recruited from the Low Vision Clinic of the Center for Ophthalmology, University of Tuebingen, Germany, from April 19, 2012, to June 6, 2012. All participants gave their written informed consent.

Inclusion criteria were diagnosis of stable maculopathy with central scotoma, magnification requirement of at least 2-fold in the better eye, and German as native language. Exclusion criteria were any additional eye disease that might affect reading; cognitive or reading impairment by medication, or any other disease.

Magnification requirement was assessed by critical print size using the Zeiss chart at 25 cm distance and +4 diopters presbyopia correction. Best-corrected far visual acuity was assessed by ETDRS charts for each eye. All participants underwent a slit-lamp examination and fundoscopy.

Radner Reading Charts

The German Radner Reading Charts [5, 6] are comparable to the MNread text charts in English [1]. It is a standardized text of 1 sentence with 14 words in Arial font. The level of difficulty corresponds to third-grade school children (8–9 years old). In each sentence, the length of words, their positions, and the syntactic complexity are the same. This test is provided in different print sizes to determine critical print size. There are 3 different samples of each print size in the version of 1998 (text #1, #2, and #3). We used 3 Radner sentences that are provided in paper print size (1 M) in the 1998 edition.

IREST Reading Charts

The IREST [8] are paragraphs of common paper print size (1.25 M) of about 132 words in Times New Roman font. The texts were developed to examine RS under everyday reading conditions, like reading paragraphs in the newspaper while using the individually adjusted magnifying visual aid. The difficulty of the texts is comparable to the level of sixth-grade school children (12–13 years of age). The texts are standardized regarding difficulty, length, and linguistic complexity [12, 29]. We chose 3 text samples of the same performance category (#3, #6, and #10).

Reading Procedure

All participants read 6 texts binocularly with best-corrected near vision and magnification on a closed-circuit image-enlarging TV system (CCTV, Reinecker Videomatic RP V4.0) at a distance of 33 cm with the head stabilized by a forehead rest. The 3 Radner texts and 3 IREST were read in randomized but alternating order (Radner text first, then IREST, or vice versa). The magnification was chosen according to the individual magnification requirement determined by the Zeiss chart with an addition of 20% to display the print above threshold. In participants with magnification requirement >4-fold, the texts on the viewing table had to be moved, which was always done by the same examiner to provide equal conditions for all patients, as not all of the participants were familiar with the handling, and to minimize variations of manual moving speed. The examiner made sure that moving speed was adequate and did not influence RS.

In accordance with the standards of the developers of the tests, the participants were instructed to read each text aloud and fluently without correcting mistakes. Reading time was measured by stopwatch, reading errors were counted for each text, and RS was calculated as number of correctly read words per minute. Each text was initially covered by a sheet of paper. The participants started to read as soon as the sheet was removed. The stopwatch was started, when the first word was spoken and was stopped right after the last word was pronounced. Before actual testing, participants were presented with 1 similar text as an example to demonstrate the test-

ing conditions. As the texts should be unknown to the participants, each text was read only once. Repeated measurements were based on the equivalence of the texts within each text length (set of 3 single sentences or 3 paragraphs). RS was always measured by the same examiner (T.M.). The participants were given time to recover between texts, if needed.

Statistical Analysis

Assessment of normality of data distribution was performed by the Shapiro-Wilk test and graphical Q-Q plots. Correlations were calculated using the Pearson correlation coefficients for RS between all pairs of 6 texts. To compare the RS between the reading charts, dependent pairwise *t* tests were performed. The pairwise *t* tests were used, because primarily differences between 2 charts independently were focus of this study, not differences between 6 charts, that is, which chart performs best or worst. Therefore, the null hypothesis in this case was whether the distribution of RS between 2 individual charts was different or not, regardless of the other pairs of charts. In addition, a 1-way repeated measures ANOVA was conducted for RS measurements with all 6 charts. The assumption of sphericity, that is, if the differences between the levels of the within-subjects factor have equal variances, was tested with Mauchly's test of sphericity. The sample effect size is reported based on the within-subjects factor variability partial eta squared (η^2).

When comparing methods, it is advisable not to rely on correlation coefficients and *p* values of test results alone, because generally they do not allow clinically relevant interpretations of the results. Thus, they can serve only as an indication but not as a proof for the conformity of 2 methods. Instead, Bland and Altman [30] suggest to estimate repeatability and agreement to express the variability between different methods [31]. Agreement is defined as a way to quantify how different tests can produce identical results [32]. Agreement, the mean and the difference in RS for each pair of charts were calculated and plotted according to Bland and Altman. The mean difference of the RS between 2 charts in all subjects (mdiff) was calculated, which is also defined as "bias" [30]. In addition, the SD, the standard error of the mean (SEM), and the 95% confidence intervals (CI) for mdiff were calculated. The 95% limits of agreement, into which 95% of the differences fall, were determined as $\text{mdiff} \pm 1.96 \times \text{SD}$. The lower the range of the 2 limits, the higher the agreement between the 2 charts (shown in Table 2). Additionally, CIs for the limits of agreement were calculated using the exact 2-sided tolerance approach as proposed by Carkeet [33]. The CIs for the limits of agreement are the approximation of 2 boundaries: the inner and outer CIs. The outer CI represents a 97.5% probability that at least 95% of the population is contained within that interval. The inner CI represents a 2.5% probability that at least 95% of the population is contained within that interval [33]. Reproducibility refers to repeated measurements if there is a variation in 1 or more conditions (either the same experiment several times or different procedures to measure the same parameter). In this study, the variation of conditions consisted of the variation of texts (Radner 1, 2, and 3 vs. IReST 3, 6, and 10) and text length. Hence, the reproducibility was measured as inter-chart reliability and agreement according to the Bland-Altman analysis [34]. Reproducibility is calculated with the same method as repeatability [30, 35]. Reproducibility was determined as $1.96 \times \sqrt{2} \times \text{within-subject SD of RS between 2 charts}$. This value is called the coefficient of repeatability [30].

Repeatability refers to measurements performed with identical conditions, that is, with all factors kept constant. Repeatability refers to the term test-retest reliability. Measurements of repeatability could not be applied, because this study did not intend to conduct any follow-up measurements. None of the texts was presented more than once, since a different chart for each measurement was used.

The level of significance was set to $\alpha = 0.05$. Statistical analyses were performed using statistical software IBM SPSS Statistics for Windows (version 25.0. Armonk, NY: IBM Corp. Released 2017).

Results

Of the 25 participants (20 females) with a central scotoma due to maculopathy, 24 had stable AMD (15 atrophic and 9 neovascular) and 1 had Best vitelliform macular dystrophy. Their mean age was 77.8 years ($\text{SD} \pm 9.9$, range 49–91 years). Regarding their reading habits, 20 of the participants were reading at least half an hour per day. Twelve participants had lower secondary education, 9 post-secondary non-tertiary education, and 4 tertiary education. Mean binocular visual acuity was 0.22 decimal ± 0.14 SD (0.65 logMAR ± 0.85 SD). The median magnification requirement was 3-fold (range 2- to 12-fold, IQR 2.25–6.13).

Mean RS for the Radner charts (#1, 2, and 3) was 92.6 wpm ± 37.2 SD, and for the IReST charts (# 3, 6, and 10), 95.0 wpm ± 37.8 SD (shown in Table 1). RS for each type of material were not statistically different from each other, except for the Radner text #3, which differed statistically from the RS of all other texts (pairwise dependent *t* tests, Table 2, second last column).

A 1-way repeated measures ANOVA elicited statistically significant changes in RS between the charts, $F(2.26, 54.18) = 4.54$, $p = 0.012$, partial $\eta^2 = 0.16$. Post hoc analysis revealed that RS was higher only for Radner #3 (98.074 ± 8.33 wpm) compared with Radner 2 (82.941 ± 7.468 wpm), a statistically significant difference of 15.132 (95% CI, 3.016–27.249) wpm, $p < 0.007$. There was a high correlation between the RS of all charts: for sentences, $r = 0.81$ – 0.90 , and for paragraphs, $r = 0.98$ (for details, see Table 2, last column).

Reproducibility: The inter-chart reliability is given in Table 3. The coefficient of repeatability between 2 charts for the IReST charts (12.9 wpm) was smaller compared with the value for the Radner charts (36.4 wpm). The coefficient of repeatability for the Radner charts was smallest between #1 and #2 and largest between #1 and #3. For the IReST charts, the smallest coefficient of repeatability was found between #3 and #10, the largest between #6 and #10. Further details are shown in Table 3.

Table 1. Descriptive data of reading speed and reading time of all charts

	Mean RS (in wpm)	SD	SEM	95% CI		Median reading time in sec (IQR)
				lower	upper	
Radner 1	96.90	37.90	7.58	81.25	112.54	7.56 (6.70–11.22)
Radner 2	98.07	41.67	8.33	80.87	115.27	7.65 (6.51–13.27)
Radner 3	82.94	37.34	7.47	67.53	98.35	8.47 (6.79–14.15)
IReST 3	95.42	38.69	7.74	79.45	111.39	75.02 (67.71–124.87)
IReST 6	93.90	39.36	7.87	77.65	110.14	76.40 (61.90–131.51)
IReST 10	95.57	35.90	7.18	80.75	110.39	75.39 (62.35–123.60)
Radner	92.64	37.18	7.44	77.29	107.98	8.09 (6.70–13.37)
IReST	94.96	37.75	7.55	79.38	110.55	76.28 (62.71–125.05)

RS, reading speed; wpm, words per minute; SD, standard deviation; SEM, standard error of the mean; CI, confidence interval; IQR, interquartile range; IReST, International Reading Speed Texts.

Agreement between the tests is shown in Table 2. Bland-Altman plots are shown in Figure 1. Agreements between the IReST charts were higher and closer together (28.97–31.01 wpm) than between the Radner charts (72.88–91.11 wpm). Between sentences and paragraphs, the highest agreement was found between Radner #1 and IReST #10 and the lowest agreement between Radner #3 and IReST #6.

An influence of the examiner could not be observed, as pointed out in Figure 2. The spread of RS was similar between the various magnification needs, and a nonlinear change between a magnification need of 2-fold (without moving of the viewing table) and 4-fold (with moving of the viewing table by the examiner) was not eminent.

Discussion

This study was conducted to compare RS of paragraphs and single sentences in maculopathy patients to investigate the repeatability of RS measurements and to evaluate which text type should be used in a clinical setting or in studies on participants with maculopathy. This type of investigation on maculopathy patients had been missing before [36].

The first result was that RS for IReST paragraphs are well comparable to those for the sentences from the Radner Reading Charts. Second, the current study has shown better agreement and better reproducibility for paragraphs compared with single sentences (shown in Table 2; Fig. 1) as in a previous study on German normally sighted older adult participants [10]. This result also cor-

responds with a study using Dutch texts on normally sighted participants [11].

For a single measurement of RS, or for assessing critical print size, the Radner charts are suitable as a quick test in patients with maculopathy. This was also suggested by Rubin [17], which the present study confirmed. For repeated measurements, paragraphs show less variability, also for patients with maculopathy. Rubin [17] discusses the use of single sentence reading tests (MNread) for measuring RS and critical print size and the use of IReST paragraphs for measuring RS with available low vision aids.

Comparing the results between the participants with maculopathy in this study with the normally sighted subjects in the previous study by Altpeter et al. [10], the correlation coefficients between single sentences were higher for the participants with maculopathy than for the normally sighted older adult group, but lower than those between paragraphs. The correlation coefficients for paragraphs in the normally sighted older participants [10] were very similar to the ones in participants with maculopathy. However, it should be considered that correlation coefficients can serve only as an indication but not as proof of conformity of 2 methods [30].

Several factors may influence reading performance and RS in patients with maculopathy: Magnification requirement plays an important role. The RS for the maculopathy participants with a relative high magnification need in the present study agrees with other studies. The higher the magnification requirement, the lower the RS, which was also confirmed in a recent study on AMD patients [22]. Linked to the magnification requirement is

Table 2. Agreement on reading speed between texts according to the Bland-Altman analysis, *t* tests, and Pearson correlation coefficients

	Mdiff ("bias")	SD	SEM	95% CI for mdiff		LoA		Confidence limits for LoA				<i>t</i> test	Correlation (Pearson <i>r</i>)	
				lower	upper	lower	upper	outer (97.5%)		inner (2.5%)				
								lower	upper	lower	upper			
Radner 1–Radner 2	–1.18	19.33	3.87	–9.16	6.80	–39.07	36.72	75.79	–55.11	52.75	–31.21	28.86	<i>p</i> > 0.05	0.89**
Radner 1–Radner 3	13.95	23.24	4.65	4.36	23.55	–31.60	59.51	91.11	–50.87	78.78	–22.15	50.06	<i>t</i> (24) = 3.00, <i>p</i> = 0.006 , <i>d</i> = 0.60	0.81**
Radner 2–Radner 3	15.13	18.59	3.72	7.46	22.81	–21.31	51.57	72.88	–36.73	66.99	–13.73	44.01	<i>t</i> (24) = 4.07, <i>p</i> = 0.00044 , <i>d</i> = 0.81	0.90**
Radner 1–IReST 3	1.47	13.54	2.71	–4.11	7.06	–25.06	28.01	53.07	–36.29	39.23	–19.55	22.50	<i>p</i> > 0.05	0.94**
Radner 1–IReST 6	3.00	14.85	2.97	–3.13	9.13	–26.12	32.11	58.23	–38.44	44.43	–20.08	26.07	<i>p</i> > 0.05	0.93**
Radner 1–IReST 10	1.32	12.70	2.54	–3.92	6.57	–23.58	26.22	49.79	–34.11	36.75	–18.41	21.05	<i>p</i> > 0.05	0.94**
Radner 2–IReST 3	2.65	20.42	4.08	–5.78	11.08	–37.37	42.67	80.04	–54.30	59.60	–29.06	34.37	<i>p</i> > 0.05	0.87**
Radner 2–IReST 6	4.18	23.42	4.68	–5.49	13.85	–41.73	50.08	91.82	–61.16	69.51	–32.21	40.56	<i>p</i> > 0.05	0.83**
Radner 2–IReST 10	2.50	17.73	3.55	–4.82	9.82	–32.24	37.24	69.48	–46.94	51.94	–25.03	30.03	<i>p</i> > 0.05	0.91**
Radner 3–IReST 3	–12.48	25.05	5.01	–22.82	–2.14	–61.57	36.61	98.19	–82.35	57.39	–51.39	26.43	<i>t</i> (24) = –2.49, <i>p</i> = 0.02 , <i>d</i> = 0.50	0.78**
Radner 3–IReST 6	–10.96	26.00	5.20	–21.69	–0.22	–61.92	40.01	101.92	–83.48	61.57	–51.34	29.43	<i>t</i> (24) = –2.11, <i>p</i> = 0.046 , <i>d</i> = 0.42	0.77**
Radner 3–IReST 10	–12.63	20.47	4.09	–21.08	–4.18	–52.75	27.48	80.23	–69.72	44.46	–44.43	19.16	<i>t</i> (24) = –3.086, <i>p</i> = 0.005 , <i>d</i> = 0.62	0.85**
IReST 3–IReST 6	1.53	7.39	1.48	–1.53	4.58	–12.96	16.01	28.97	–19.09	22.14	–9.95	13.01	<i>p</i> > 0.05	0.98**
IReST 3–IReST 10	–0.15	7.79	1.56	–3.37	3.07	–15.43	15.12	30.55	–21.89	21.59	–12.26	11.95	<i>p</i> > 0.05	0.98**
IReST 6–IReST 10	–1.68	7.91	1.58	–4.94	1.59	–17.18	13.83	31.01	–23.75	20.39	–13.97	10.61	<i>p</i> > 0.05	0.98**

According to Bland-Altman, the differences between 2 measurements should lie within a range of agreement, where 95% of all differences fall (LoA). The range is defined as lower = mdiff – 1.96 × SD and upper = mdiff + 1.96 × SD. For the LoA, outer and inner confidence intervals are provided [33]. The level of significance is 0.05, if not otherwise stated.

Mdiff, mean difference of reading speed between 2 charts (= "bias"); SD, standard deviation; SEM, standard error of the mean; CI, confidence interval; LoA, limits of agreement; *t* test, paired dependent *t* tests between reading speeds of paired reading charts; *p*, probability value; correlation, the Pearson product-moment correlation coefficient *r*; *d*, Cohen's measure of effect size; outer, the outer 97.5% confidence interval for the LoA, where 95% of the population fall into; inner, the inner 2.5% confidence interval for the LoA, where 95% of the population fall into; IReST, International Reading Speed Tests. ** The correlation is significant at the level of 0.01 (2-sided).

Table 3. Inter-chart reliability

	Mean within-subject SD	Repeatability coefficient	95% CI for repeatability coefficient	
			lower	upper
Radner 1–Radner 2	9.37	25.97	14.75	37.19
Radner 1–Radner 3	13.66	37.86	22.64	53.09
Radner 2–Radner 3	12.88	35.70	23.20	48.20
IReST 3–IReST 6	4.32	11.98	8.53	15.42
IReST 3–IReST 10	3.81	10.55	6.07	15.02
IReST 6–IReST 10	4.34	12.03	7.88	16.18
Radner	13.15	36.44	24.78	48.09
IReST	4.65	12.90	9.66	16.13

Mean within participant SD, the mean within participant standard deviation between reading speed measurements in words per minute; repeatability coefficient, $1.96 \times \sqrt{2} \times$ within-subject standard deviation with 95% CI; CI, confidence interval; IReST, International Reading Speed Texts.

Fig. 1. Bland-Altman plots for reading speed between Radner and IReST charts. **a–d** Examples of Bland-Altman plots for Radner (**a**, **b**) and IReST reading charts (**c**, **d**). Abscissa: Mean reading speed of 2 charts in wpm. Ordinate: Difference in reading speed between 2 charts in wpm. Dashed line: The mean difference (bias) with confidence interval for the bias (error bar). Upper and lower solid lines represent the 95% limits of agreement (= bias $\pm$ 1.96 $\times$ SD). The

outer confidence interval represents a 97.5% probability that at least 95% of the population is contained within that interval. The inner confidence interval represents a 2.5% probability that at least 95% of the population is contained within that interval (solid lines on the right). RS, reading speed; diff., difference; wpm, words per minute; SD, standard deviation; IReST, International Reading Speed Texts.

Fig. 2. Scatter diagram for the mean reading speed plotted against the magnification need. Mean reading speed of Radner texts is found on the ordinate to the left, mean reading speed of IReST on the ordinate to the right, and the magnification need on the abscissa. Black dots represent the mean reading speed for the 3 Radner texts of 1 participant, and grey triangles show the mean reading speed for the 3 IReST of 1 participant. IReST, International Reading Speed Texts.

the severity and stage of the disease: the larger the central scotoma, the more eccentric the PRL and the higher the magnification requirement. Patients with early and intermediate AMD showed only slightly reduced RS: Pondorfer et al. [37] reported 159.4 (± 21.4) wpm for early and 153.8 (± 29.2) wpm for intermediate AMD measured with IReST. Their RS were much closer to the RS of normally sighted older adults (167 wpm) in the study by Altpeter et al. [10], showing as well the major role of age for RS [7, 18]. In addition, the position of the central scotoma in patients with maculopathies correlates with the probability of misinterpreting letters at the beginning and the end of words [3].

The method of magnification may also play a role. The participants in this study were all reading on a CCVT with their individual magnification in order to receive equal conditions for all participants without additional influence of different magnifying visual aids. If the effect of different magnifying systems to RS is the subject of a study, a crucial point is the amount of the individual and exactly determined magnification requirement and the task. If magnification requirement is more than 6-fold, the CCVT method is more comfortable and therefore probably less fatiguing because of the larger overview and better ergonomic posture compared to optical aids or hand-held electronic devices.

New electronic devices have the advantage that they are easily accessible and not stigmatizing. Recent studies on the use of smartphones as spot-reading magnifiers compared to a portable video magnifier [38] showed comparable results between the 2 methods. The compar-

ison between smartphones and CCTV devices or tablets showed good acceptance, especially in people, who were already familiar with these devices, and demonstrated even better performance in experienced patients [39, 40]. In these studies, the participants were low vision patients with different diagnoses, only part of whom had a central scotoma, which can influence the reading deficit in different ways. It is expected that low vision patients will use modern electronic devices more frequently in the future, so that an individual consultation about their optimal use for specific tasks (e.g., spot reading, longer text reading, navigation, and orientation) will be desirable.

Furthermore, greater variability is to be expected if timing is performed by stopwatch. Specifically, reading time is underestimated by 0.3 s if a stopwatch is used [41]. Xu et al. [41] used a computer-based reading test with short sentences of 12 words and found that the variability of measurement by stopwatch by the same examiner for every sentence or paragraph is negligible. This has been confirmed by a recent study by Radner et al. [42], in which an automated computer program for measuring reading time was compared with stopwatch measurements.

In patients, reading ability might also be influenced by the font of the text. Tarita-Nistor et al. [43] examined patients with age-related macular degeneration and found a poorer reading acuity for text set in Arial than in Times New Roman: in this regard, the Radner texts (Arial) may have a slight disadvantage compared with IReST (Times New Roman). However, Rubin et al. [44] showed in people with mild-to-moderate visual impairment that the typeface had little influence on RS.

While the current study only included participants with maculopathy in order to work with a homogeneous group, several studies measured RS in participants with different eye diseases: Burggraaff et al. [45] measured RS with the Dutch version of the Radner charts in 38 low vision patients with different diagnoses (50% with maculopathy), which resulted in a mean RS of 123 wpm (SD $\pm$ 30.9). They found a lower inter-chart and test-retest reliability for the Radner test compared with more homogenous participant groups as in the study of Stifter et al. [46], which examined well-defined groups of participants with maculopathy with the Radner test.

In English-speaking countries, MNread is the most frequently used test for measuring RS and critical print size [1], like the Radner charts are in German-speaking countries. The single sentence structure of MNread is comparable to that of the Radner charts. The syntactical complexity is slightly more homogenous in Radner texts with 14 words of the same length and position than in MNread. The MNread sentences consist of 3 lines with 10–14 words (60 characters), and the length and positions of words vary slightly more than in the Radner texts.

Patel et al. [47] examined the RS in 59 AMD patients with MNread sentences and found a higher test-retest variability than Subramanian and Pardhan [48], who tested the repeatability of MNread in 27 low vision patients mostly due to maculopathy. In both studies, RS measured by MNread sentences were faster than in this study, which was most likely caused by the higher magnification requirements of the participants in this study.

There are several reasons why the measurements for paragraphs were more homogenous: paragraphs contain more sentences, so that the result is always an average of a higher number of samples, which results in less noise and variance. This is the predominant reason why the IReST paragraphs were developed [15, 16]. The lower variability between the IReST paragraphs could have been partly caused by the lower percentage of error of the reading time assessment in seconds, which is smaller for longer texts: paragraphs are less susceptible to inaccuracies in timing [7]. Additionally, the linguistic equivalence of the IReST paragraphs contributed to the lower variability between RS for the IReST paragraphs. Linguistic factors influence RS, for example, in regard to predictability of the text, orthographic regularity, and word length and word frequency [12, 14, 29, 36]. Another advantage of paragraphs is that they can provide some information about fluency, fatigue, and mistakes. On the other hand, the reading time for a paragraph is longer compared to a single sentence, which

might be seen as a disadvantage in a time-pressured clinical setting. However, a goal for developing the IReST paragraphs was the balance between the advantage of a paragraph, compared with a single sentence and the disadvantage of a long reading time for a longer text, as they are used for longer silent reading to assess fatigue [16]. This shows that, dependent on the question, different reading tests can be indicated, and which combination can achieve especially conclusive information.

Conclusion

In participants with maculopathy, no significant difference in mean RS measured by single sentences or paragraphs was shown, as in a previously examined normally sighted older group [10]. However, like normally sighted participants, patients with maculopathy also showed a higher variability of RS between single sentences than paragraphs. We conclude that single sentences are well suited for measuring RS and critical print size. For repeated measurements of RS, however, whole paragraphs are preferable because of the lower variability of RS.

Acknowledgement

We thank Manfred MacKeben, PhD, The Smith-Kettlewell Eye Research Institute, San Francisco, CA, for helpful discussions and editing the manuscript.

Statement of Ethics

All patients gave their informed written consent. Approval and protocol for this study were obtained from the Ethics Committee of the Medical Faculty, University of Tübingen, Germany, on April 18, 2012. All procedures complied with the tenets of the Declaration of Helsinki.

Conflict of Interest Statement

C.K., T.M., and E.K.A. declare that they have no conflicts of interest. At the time of the study, and in the previous 3 years, the Center of Ophthalmology, University of Tübingen, received royalties for the sale of the IReST charts. From June 2020, royalties have gone to Odilia Vision® GmbH, a German limited liability company, of which Stephan Kuester-Gruber and Susanne Trauzettel-Klosinski are shareholders.

Author Contributions

C.K. drafted the manuscript and revised it critically. T.M. was responsible for the acquisition of the data, for drafting and revising the content, and approved the final version. E.K.A. contributed to the concept of the study, contributed to data acquisition, drafted and revised the content, and approved the final version. S.T.-K.

developed the concept of the study, supervised the study, and wrote the final version of the manuscript. S.K.-G. performed the statistical calculations and revised the manuscript critically for intellectual content. All authors agree to be accountable for all aspects of the work by assuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated.

References

- Legge GE, Ross JA, Luebker A, LaMay JM. Psychophysics of reading. VIII. The Minnesota low-vision reading test. *Optom Vis Sci.* 1989;66(12):843–53.
- Colenbrander A, Fletcher DC. Low vision rehabilitation: visual acuity measurement in the low vision range. *J Ophthalmic Nurs Technol.* 1992;11(2):62–9.
- MacKeben M, Nair UK, Walker LL, Fletcher DC. Random word recognition chart helps scotoma assessment in low vision. *Optom Vis Sci.* 2015;92(4):421–8.
- Eisenbarth W, Pado U, Schriever S, Schötschel D, Feucht N, MacKeben M. [Localization of scotomas in AMD by reading test: random series of words in standardized format]. *Ophthalmologe.* 2016;113(9):754–62.
- Radner W, Obermayer W, Richter-Mueksch S, Willinger U, Velikay-Parel M, Eisenwort B. The validity and reliability of short German sentences for measuring reading speed. *Graefes Arch Clin Exp Ophthalmol.* 2002;240(6):461–7.
- Radner W, Willinger U, Obermayer W, Mudrich C, Velikay-Parel M, Eisenwort B. [A new reading chart for simultaneous determination of reading vision and reading speed]. *Klin Monbl Augenheilkd.* 1998;213(3):174–81.
- Hahn GA, Penka D, Gehrlich C, Messias A, Weismann M, Hyvärinen L, et al. New standardised texts for assessing reading performance in four European languages. *Br J Ophthalmol.* 2006;90(4):480–4.
- Trauzettel-Klosinski S, Dietz K, Group IRS. Standardized assessment of reading performance: the new international reading speed texts IReST. *Invest Ophthalmol Vis Sci.* 2012;53(9):5452–61.
- Radner W, Radner S, Diendorfer G. A new principle for the standardization of long paragraphs for reading speed analysis. *Graefes Arch Clin Exp Ophthalmol.* 2016;254(1):177–84.
- Altpeter EK, Marx T, Nguyen NX, Naumann A, Trauzettel-Klosinski S. Measurement of reading speed with standardized texts: a comparison of single sentences and paragraphs. *Graefes Arch Clin Exp Ophthalmol.* 2015;253(8):1369–75.
- Brussee T, van Nispen RM, Klerkx EM, Knol DL, van Rens GH. Comparison of reading performance tests concerning difficulty of sentences and paragraphs and their reliability. *Ophthalmic Physiol Opt.* 2015;35(3):324–35.
- Gibson E. Linguistic complexity: locality of syntactic dependencies. *Cognition.* 1998;68(1):1–76.
- Gleni A, Ktistakis E, Tsilimbaris MK, Simos P, Trauzettel-Klosinski S, Plainis S. Assessing variability in reading performance with the New Greek standardized reading speed texts (IReST). *Optom Vis Sci.* 2019;96(10):761–7.
- Duerwaechter U, Sokolov AN, Reinhard J, Klosinski G, Trauzettel-Klosinski S. Word length and word frequency affect eye movements in dyslexic children reading in a regular (German) orthography. *Ann Dyslexia.* 2010;60(1):86–101.
- Ramulu PY, Swenor BK, Jefferys JL, Friedman DS, Rubin GS. Difficulty with out-loud and silent reading in glaucoma. *Invest Ophthalmol Vis Sci.* 2013;54(1):666–72.
- Ramulu PY, Swenor BK, Jefferys JL, Rubin GS. Description and validation of a test to evaluate sustained silent reading. *Invest Ophthalmol Vis Sci.* 2013;54(1):673–80.
- Rubin GS. Measuring reading performance. *Vision Res.* 2013;90:43–51.
- Legge GE, Ross JA, Isenberg LM, LaMay JM. Psychophysics of reading. Clinical predictors of low-vision reading speed. *Invest Ophthalmol Vis Sci.* 1992;33(3):677–87.
- Trauzettel-klosinski S, Tornow R-P. Fixation behavior and reading ability in macular scotoma: assessed by Tuebingen manual perimetry and scanning laser ophthalmoscopy. *Neuro-Ophthalmology.* 1996;16(4):241–53.
- Trauzettel-Klosinski S, Teschner C, Tornow R-P, Zrenner E. Reading strategies in normal subjects and in patients with macular scotoma-assessed by two new methods of registration. *Neuro-Ophthalmology.* 1994;14(1):15–30.
- Calabrese A, Bernard JB, Hoffart L, Faure G, Barouch F, Conrath J, et al. Wet versus dry age-related macular degeneration in patients with central field loss: different effects on maximum reading speed. *Invest Ophthalmol Vis Sci.* 2011;52(5):2417–24.
- Kaltenegger K, Kuester S, Altpeter-Ott E, Eschweiler GW, Cordey A, Ivanov IV, et al. Effects of home reading training on reading and quality of life in AMD – a randomized and controlled study. *Graefes Arch Clin Exp Ophthalmol.* 2019;257(7):1499–512.
- Crossland MD, Culham LE, Rubin GS. Fixation stability and reading speed in patients with newly developed macular disease. *Ophthalmic Physiol Opt.* 2004;24(4):327–33.
- Crossland MD, Rubin GS. Eye movements and reading in macular disease: further support for the shrinking perceptual span hypothesis. *Vision Res.* 2006;46(4):590–7.
- Calabrese A, Bernard JB, Faure G, Hoffart L, Castet E. Eye movements and reading speed in macular disease: the shrinking perceptual span hypothesis requires and is supported by a mediation analysis. *Invest Ophthalmol Vis Sci.* 2014;55(6):3638–45.
- Trauzettel-Klosinski S. Rehabilitation for visual disorders. *J Neuroophthalmol.* 2010;30(1):73–84.
- Trauzettel-Klosinski S. Reading disorders due to visual field defects: a neuro-ophthalmological view. *Neuro-Ophthalmology.* 2009;27(1–3):79–90.
- Marx T. *Vergleich zweier deutschsprachiger, standardisierter Lesetafeln zur Bestimmung der Lesegeschwindigkeit [Inaugural Dissertation]*. Tübingen: Eberhard Karls Universität; 2015.
- Gibson E. The dependency locality theory: a distance-based theory of linguistic complexity. In: Marantz A, Miyashita Y, O’Neil W, editors. *Image, Language, Brain*. Cambridge, MA: MIT Press; 2000. p. 95–126.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8(2):135–60.
- Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol.* 2008;31(4):466–75.
- Berchold A. Test-retest: agreement or reliability? *Methodol Innovations.* 2016;9:2059799116672875.
- Carkeet A. A review of the use of confidence intervals for Bland-Altman limits of agreement in optometry and vision science. *Optom Vis Sci.* 2020;97(1):3–8.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307–10.
- McAlinden C, Khadka J, Pesudovs K. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic Physiol Opt.* 2011;31(4):330–8.
- Brussee T, van Nispen RM, van Rens GH. Measurement properties of continuous text reading performance tests. *Ophthalmic Physiol Opt.* 2014;34(6):636–57.

- 37 Pondorfer SG, Terheyden JH, Heinemann M, Wintergerst MWM, Holz FG, Finger RP. Association of vision-related quality of life with visual function in age-related macular degeneration. *Sci Rep*. 2019;9(1):15326.
- 38 Wittich W, Jarry J, Morrice E, Johnson A. Effectiveness of the Apple iPad as a spot-reading magnifier. *Optom Vis Sci*. 2018;95(9):704–10.
- 39 Martiniello N, Eisenbarth W, Lehane C, Johnson A, Wittich W. Exploring the use of smartphones and tablets among people with visual impairments: are mainstream devices replacing the use of traditional visual aids? *Assist Technol*. 2019:1–12.
- 40 Morrice E, Johnson AP, Marinier JA, Wittich W. Assessment of the Apple iPad as a low-vision reading aid. *Eye*. 2017;31(6):865–71.
- 41 Xu R, Bradley A. IURead: a new computer-based reading test. *Ophthalmic Physiol Opt*. 2015;35:500–513.
- 42 Radner W, Diendorfer G, Kainrath B, Kollmitzer C. The accuracy of reading speed measurement by stopwatch versus measurement with an automated computer program (rad-rd(c)). *Acta Ophthalmol*. 2017;95(2):211–6.
- 43 Tarita-Nistor L, Lam D, Brent MH, Steinbach MJ, González EG. Courier: a better font for reading with age-related macular degeneration. *Can J Ophthalmol*. 2013;48(1):56–62.
- 44 Rubin GS, Feely M, Perera S, Ekstrom K, Williamson E. The effect of font and line width on reading speed in people with mild to moderate vision loss. *Ophthalmic Physiol Opt*. 2006;26(6):545–54.
- 45 Burggraaff MC, van Nispen RM, Hoek S, Knol DL, van Rens GH. Feasibility of the Radner reading charts in low-vision patients. *Graefes Arch Clin Exp Ophthalmol*. 2010;248(11):1631–7.
- 46 Stifter E, König F, Lang T, Bauer P, Richter-Müksch S, Velikay-Parel M, et al. Reliability of a standardized reading chart system: variance component analysis, test-retest and inter-chart reliability. *Graefes Arch Clin Exp Ophthalmol*. 2004;242(1):31–9.
- 47 Patel PJ, Chen FK, Da Cruz L, Rubin GS, Tufail A. Test-retest variability of reading performance metrics using MNREAD in patients with age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2011;52(6):3854–9.
- 48 Subramanian A, Pardhan S. Repeatability of reading ability indices in subjects with impaired vision. *Invest Ophthalmol Vis Sci*. 2009;50(8):3643–7.