

Ophthalmic Research

Ophthalmic Res , DOI: 10.1159/000520029

Received: March 2, 2021

Accepted: September 24, 2021

Published online: October 5, 2021

Methodical bias, incorrect use of the RADNER Reading Charts and - avoiding misleading reporting in Visual Acuity Measurements Kortuem C, Marx T, Altpeter E, Trauzettel-Klosinski S, Kuester-Gruber S. Comparing reading speeds for reading standardized single sentences and paragraphs in patients with maculopathy. Ophthalmic Res. 2021;64(3):512-522.

Radner W, Maaijwee K, de Smet MD, Benesch T, Ettl A

ISSN: 0030-3747 (Print), eISSN: 1423-0259 (Online)

<https://www.karger.com/ORE>

Ophthalmic Research

Disclaimer:

Accepted, unedited article not yet assigned to an issue. The statements, opinions and data contained in this publication are solely those of the individual authors and contributors and not of the publisher and the editor(s). The publisher and the editor(s) disclaim responsibility for any injury to persons or property resulting from any ideas, methods, instructions or products referred to the content.

Copyright:

This article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC) (<http://www.karger.com/Services/OpenAccessLicense>). Usage and distribution for commercial purposes requires written permission.

Accepted Manuscript

Methodical bias, incorrect use of the RADNER Reading Charts and - avoiding misleading reporting in Visual Acuity Measurements in Kortuem C, Marx T, Altpeter E, Trauzettel-Klosinski S, Kuester-Gruber S. Comparing reading speeds for reading standardized single sentences and paragraphs in patients with maculopathy. Ophthalmic Res. 2021;64(3):512-522.

Wolfgang Radner^{1,5}, Kristel Maaijwee², Marc D. de Smet³, Thomas Benesch⁴

Armin Ettl⁵

Wolfgang Radner, MD., Prof., ¹Austrian Academy of Ophthalmology, Mollgasse 11, 1180 Vienna, Austria;
⁵Universitätsklinikum St. Pölten, Klinische Abteilung für Augenheilkunde und Orbitachirurgie; Karl Landsteiner Medical University, Krems; wolfgang.radner@chello.at

Kristel Maaijwee, MD., PhD, Kristel Maaijwee, MD., PhD, ²HagaZiekenhuis, Department of Ophthalmology, Els-Borst Eilersplein 275, 2545 AA The Hague, the Netherlands; kmaaijwee@hotmail.com

Marc D. de Smet MDCM, PhD, MIOS, ³Lausanne Switzerland; Dept of Ophthalmology, University of Leiden, the Netherlands mddesmet1@mac.com

Thomas Benesch, PhD. PhD, Prof. ⁴Baumgasse 16, 1030 Vienna, Austria,
thom@s-benesch.com

Armin Ettl, MD., PhD, Univ.-Doz. ⁵Universitätsklinikum St. Pölten, Klinische Abteilung für Augenheilkunde und Orbitachirurgie; Karl Landsteiner Medical University, Krems, Austria, armin.Ettl@stpoelten.lknoe.at

Short title: Incorrect use of the RADNER Reading Charts

Word count: 2610

Number of figures: 1

Key words: reading charts, reading speed, repeated measures, reading acuity, reading performance

Address correspondence to:

Wolfgang Radner, MD, Prof.

Austrian Academy of Ophthalmology

Mollgasse 11

A- 1180 Vienna,

Austria

Phone: +43 1 5231492

Fax: +43 1 523 14 92 1

E-mail: wolfgang.radner@chello.at

Accepted Manuscript

Dear Editor-in-Chief:

We read with interest the publication of Kortuem et al. We greatly appreciate the authors' opinion that the RADNER Reading Charts provide the most carefully standardized test items (sentence optotypes) of all the reading tests in ophthalmology.

In using the RADNER Reading Charts as well as other reading charts, particularly for clinical studies, it is essential that the correct test protocol is followed; otherwise, the test results cannot be properly interpreted and compared to previous studies. This is an issue that has already been discussed in a previous publication by this group [1] and already clarified by a letter to the editor [2].

Although an ERRATUM has been published, reviewing the sources of methodological biases here and how to avoid them would be useful to readers of the journal:

Preamble

The study by Kortuem et al. is limited by the fact that in it the RADNER Reading Charts were not employed in accordance with their intended use and that the reading speed obtained with only 3 out of 24 sentences of the RADNER Reading Charts was not calculated in accordance with the statistically certified standard procedure for those charts.

Information from the data source (thesis) [3] raises the question of why the IReST was recommended for "repeated use" when almost 30% of the final number of eligible participants (n=25) were unable to finish the study, and an additional unknown number were excluded during a practice session with IReST paragraph 2.

Sources of methodological bias:

1) Incorrect data acquisition:

To be accurate, reading speeds need to be calculated according to established validated protocols. In the study of Kortuem et al., reading speeds were calculated incorrectly and not in accordance with the protocol provided for the RADNER Reading Chart. These instructions are not just meant as "recommendations" as noted in the ERRATUM; they are the result of elaborate statistical standardizations [2, 4, 5, 6, 7, 8, 9]. We have shown in a number of studies that using the RADNER Charts and the protocol as intended ensures reliable results [2, 4, 5, 6, 7, 8, 9].

Because of the standardization of the sentence optotypes used for the RADNER Charts and to ensure "lege artis" use, reading speed *must* be calculated by using the number of words per sentence (14) and the reading time per sentence (sec) in the following equation:

reading speed in words per minute (w/min) = $(14/\text{reading time in seconds}) \times 60 = 840/\text{reading time in seconds}$.

2) Incorrect visual acuity calculation:

The mean visual acuity given as "0.22 ±0.14 SD (0.65 logMAR ±0.85 SD)" in the Abstract and the Results is highly suspect.

The mean and standard deviation (SD) of various levels of visual acuity that are represented by a geometric (logarithmic) progression have to be calculated by using either -logDecimal or logMAR [10]. Using an arithmetic

calculation of a mean from decimal acuities is wrong [10], as is converting the resulting mean and SD into logMAR by using the “log button” of a calculator. An SD of ± 0.85 logMAR cannot be correct and indicates that a “mean decimal acuity and SD” have initially been erroneously calculated and then converted into logMAR.

3) Incorrect use of the RADNER Reading Charts:

The RADNER Reading Charts were designed for clinical and research use, providing a number of different reading parameters from a single examination in patients with normal to low vision [4, 5, 6, 7, 8, 9]. These reading charts offer standardized “sentence optotypes” that geometrically (logarithmically) progress in print size (Fig. 1). The sentence optotypes, developed in 12 different languages, have been standardized by reliability and validity analyses in 1,323 individuals, involving over 44,000 measurements. In addition, the reliability of the RADNER Reading Charts has been analyzed by a test–retest protocol (interval, 3 – 4 weeks). The test-retest reliability and the interchart reliability were investigated, and a variant component analysis was performed (randomized, orthogonal Latin square design) [6, 7, 8, 9]. As far as we are aware, such analyses have not been reported for the IReST.

In addition, the sentence optotypes of the RADNER Reading Charts have consistently given correlations of $r \approx 0.9$ with long paragraphs [4, 5]. Analyzing various reading parameters (such as reading acuity, mean and maximum reading speed, and critical print size, among others) has proven advantageous in evaluating specific alterations of reading performance in various eye diseases [8, 9].

The RADNER Reading Charts were never designed or intended for investigating patients by presenting just one sentence per chart [2, 4, 5, 6, 7, 8, 9].

Kortuem et al. suggest that just one RADNER sentence per chart might provide a measure of a patient’s reading speed. However, unless this approach is validated according to a protocol similar to that used to evaluate the original RADNER Reading Charts, the validity, reliability, and reproducibility of such measurements cannot be ensured. To then compare such measurements to another reading chart becomes even more questionable.

Finally, Brusse et al. [11] have shown that 6 of 10 IReST paragraphs differ significantly from the mean. This is highly relevant in clinical studies when the means of reading speeds are statistically compared.

It is well known that increasing the number of presented text passages and calculating a mean of reading speed \pm SD on a larger sample size increases the reliability of the measurements. This was the principal thought behind the design of the RADNER Reading Charts and is the reason why the use of a RADNER Chart should be limited to its intended use, as given in the instructions.

4) Statistical bias

Neither in the manuscript nor in the thesis of Marx (source of data) [3] is there information about how the randomization was performed. All participants practiced the test procedure with IReST paragraph 2, but there was no practice session prior to the use of the single sentences [3].

Exclusion of participants is explained in the thesis [3], but not in the manuscript. It is notable that 8 participants were excluded during the test procedure, and in addition an unknown number of participants were excluded from the pre-tests that involved practicing the test procedure with IReST paragraph 2. Participants were excluded if in any of the text passages they repeated parts of the text, as can happen when a reader does not find the subsequent line, or when the reader attempts to self-correct while reading the text.

These exclusions likely represent a positive selection bias for the IReST paragraphs and in favor of good responders, to the exclusion of others. Is this the explanation for the high correlation coefficient ($r=0.98$) between the IReST paragraphs? What would the correlation be without such an exclusion? Given that about 30% of the

final number of patients were excluded because they were unable to complete the test, it seems surprising to draw conclusions concerning the use of IReST for repeated measurements.

With 6 test items to be assessed, the investigation of only 25 subjects seems an unusually low number to ensure that all text paragraphs were read at the same position in the test sequence. For a study comparing reading speed obtained with 3 long paragraphs in participants with healthy eyes, a power calculation reveals the need for a minimum of 34 participants (for 80% power; Cohen's $d=0.51$ and an alpha of 5%). Cohen's d was estimated according to a previous study [12]. This calculation does not take into account the addition of single sentences, since only reading these after long paragraphs can lead to bias because of a reduced level of concentration and motivation induced by reading long texts prior to the single sentences. It is also important to consider the possibility of outliers having an undue influence on the mean and SD.

Studies in which multiple means are being compared require a multiple testing correction to be applied in their statistical analysis. The simplest and most widely used method for multiple testing correction is the Bonferroni adjustment: If a significance threshold of alpha is used, but n separate tests are performed, then the Bonferroni adjustment deems a score significant only if the corresponding P-value is less than α/n .

The authors also drew an incorrect conclusion concerning repeatability. In their study they only investigated the inter-item reliability: They state that the inter-chart reliability (coefficient of repeatability) was smaller for the paragraphs (12.9 wpm) than for the single sentences (36.4 wpm). However, subsequently, they reach an erroneous conclusion that "In patients with maculopathy, single sentences are well suited for single measurement of reading speed. 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." It is important to emphasize that they tested the inter-chart reliability, but their conclusion is about test-retest reliability (with a time lapse between two measurements), which was not investigated in their study.

5) Selection of only specific texts or sentences:

Only 3 out of 10 texts of the IReST were used.

Only 3 out of 24 sentence optotypes of the RADNER Charts were used.

Why were paragraphs 3, 6, and 10 of the IReST selected, and not 1, 2, and 3? Selection criteria are not given. Are texts 3, 6, and 10 known to be the most equivalent to each other? Brusse et al. have shown that 6 of 10 IReST paragraphs differ statistically significantly from the mean [11].

6) Bias caused by different print sizes (magnification bias):

In T. Marx's thesis (2015), from which the published data are derived, the print size of the lower case letters in the IReST is stated to be 1.7 mm [3]. The print sizes of the sentence optotypes used from the RADNER Reading Charts are 0.4 logRAD or 1.461 mm [4, 5, 13]. Thus, there is a difference between the two print sizes of about 14%, or more than half a log step; this difference is of particular relevance for low-vision patients. Although Marx [3] mentioned that print size is crucial for reading speed, he did not verify the print sizes used; for the RADNER Charts, he would only have had to calculate the required x-height for 0.4 logRAD.

In addition, the RADNER Charts give the required magnifications for a reading distance of 25 cm up to 6.25x, which are in close accord with the magnifications given in the Zeiss test that was used to estimate the required magnification in the thesis. The only difference is that the Zeiss test is printed in bold letters, and the RADNER Charts more correctly follow the geometrical sequence of print sizes [13]. This would have allowed him to use the RADNER Charts without additional magnification and, thus, more correctly in 21 of the 25 participants.

Conclusion

Measuring visual acuity and reading performance is likely to be fraught with many methodological, psychophysical, and statistical limitations and potential biases. Therefore, it is imperative that at all stages of development, very strict and rigorous protocols are followed to ensure that valid conclusions are drawn from experimental data. While it is a worthwhile endeavor to develop a simple, repeatable reading metric, it is important that each component is rigorously tested as to avoid reaching wrong conclusions because of poor experimental design or implementation. A better definition of the IReST charts and their limitations and the correct or validated use of a modified test procedure for the RADNER Reading Charts would be an appropriate starting point.

Addendum 1: The problem with using “only correctly read words” for reading speed calculation:

Reading speed calculations that are based only upon correctly read words constitute a problem because they do not differentiate between long and very short words, and many further questions remain unanswered: Are words counted as “wrong” when (a) only the singular instead of the plural is used, or (b) small words with no impact on the understanding of the content are read wrong, or (c) is a distinction made between long words of high relevance for the content and small words that are not of relevance to the content?

What decisions and how they are made by an examiner should be spelled out prior to any study to insure consistency. Furthermore, subtracting words read incorrectly causes a double punishment bias because reading errors usually reduce reading speed even when not corrected.

In contrast, for the RADNER Reading Charts, such sources of inaccuracy have been considered and avoided by establishing a “Reading Acuity Score” involving different lengths of words and by the use of a “stop criterion.”

Addendum 2: Why it is of relevance to calculate reading speed as given in the instructions:

The sentence optotypes of the RADNER Reading Charts have been statistically selected for equal difficulty and reading time [4, 5, 6, 7, 8, 9]. This effort has been made in order to facilitate their clinical and scientific use, so that these RADNER Charts can (and must be) used with “real reading speed measurements” obtained as clearly stipulated in the instructions (reading speed calculated for all 14 words of a sentence optotype). The reading speed has to be calculated by using the following equation:

reading speed in words per minute (wpm) = number of words ($n=14$) x 60 seconds divided by reading time in seconds (t) = $840 / t$.

This is of relevance because even when words are read incorrectly, their reading consumes the same proportional share of the total reading time as other words of the text consume. Thus, excluding such incorrectly read words would artificially change the ratio between the reading time and the actual number of words read. Although the number of errors per 100 words was found to be similar for long paragraphs and single sentences (0.5 errors/100 words and 0.66 errors/100 words, respectively) [12], it is evident that the proportional bias in the reading time caused by subtracting the words that are read incorrectly increases to some unknown extent with the number of errors made and with a decrease in the number of words in a text (unknown proportion bias).

Since at 85 wpm, 2 reading errors in a text of 130 words reduce the reading speed by only 1.8% and, even 10 words read incorrectly produce a reduction of only 8%, subtracting the words read incorrectly does not seem to be an adequate approach for equalizing methods between different reading tests, particularly not when it is obvious that the impact of this proportional bias is much bigger for the other test (the RADNER Charts) if the RADNER Charts are not used as intended.

Kind regards,

Wolfgang Radner, MD. Prof.

Kristel Maijwee, MD., PhD.

Marc deSmet, MDCM., PhD., MIOS, Prof.

Thomas Benesch, PhD., Prof.

Armin Ettl, MD., PhD, Prof.

Jorge Alio; MD., PhD., Prof.

Joaquim Neto Murta, MD., PhD., Prof.

Peter Vamosi, MD., PhD. Prof.

Mihnea Munteanu, MD., PhD., Prof.

Andreia Martins Rosa, MD., PhD.

Dominique Bremond-Gignac, MD., PhD., Prof.

Acknowledgements:

The authors would like to thank for their valuable input and contributions:

Prof. Jorge L Alio, MD., PhD, Chairman of Ophthalmology – Visum, Alicante, Spain;

Joaquim Neto Murta, MD., PhD., Chairman Department Ophthalmology, Centro Hospitalar Universitário Coimbra, Coimbra, Portugal;

Prof. Peter Vamosi, MD., PhD., Chairman of the Department of Ophthalmology, Péterfy Hospital, Budapest, Hungary

Prof. Mihnea Munteanu, MD., PhD.; Chairman of the Department of Ophthalmology, "Victor Babes" University of Medicine and Pharmacy, Timisoara, Romania;

Andreia Martins Rosa, MD., PhD., Faculty of Medicine, University of Coimbra, Portugal;

Prof. Dominique Bremond-Gignac, MD., PhD., Head of the Department of Ophthalmology, University Hospital Necker Enfants Malades, APHP, Paris, France.

Author Contributions:

Wolfgang Radner, Kristel Maaijwee, Marc D. de Smet, Thomas Benesch and

Armin Ettl substantially contributed on drafting, writing and critically revising the manuscript including final approval.

Funding:

No funding was received for this letter to the editor.

Conflict of interest:

W. Radner and K. Maaijwee receive royalties for the RADNER Reading Charts. All other named authors have no affiliation with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent licensing arrangements) or non-financial interest (such as personal or professional relationships, affiliations, knowledge, or beliefs) in the subject matter or materials discussed in this manuscript.

References:

1 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;253:1369-75.

2 Radner W, Maaijwee K, de Smet MD, Benesch T. Incorrect use and presentation of the RADNER Reading Charts: comment on measurement of reading speed with standardized texts: a comparison of single sentences and paragraphs, by Altpeter E, Marx T, Nguyen N, Naumann A, Trauzettel-Klosinski S. *Graefes Arch Clin Exp Ophthalmol*. 2016;254:813-5.

3 Marx T. Vergleich zweier deutschsprachiger, standardisierter Lesetafeln zur Bestimmung der Lesegeschwindigkeit [Inaugural-Dissertation]. Tübingen: Eberhard Karls Universität; 2015.

4 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;213:174-81.

5 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;240:461-7.

6 Stifter E, König F, Lang T, Bauer P, Richter-Müksch S, Velikay-Parel M, Radner W. Reliability of a standardized reading chart system: variance component analysis, test-retest and inter-chart reliability. *Graefes Arch Clin Exp Ophthalmol*. 2004;242:31-9.

7 Maaijwee K, Mulder P, Radner W, Van Meurs JC. Reliability testing of the Dutch version of the Radner Reading Charts. *Optom Vis Sci*. 2008;85:353-8.

8 Radner W. Standardization of Reading Charts: A Review of Recent Developments. *Optom Vis Sci*. 2019;96:768-79.

9 Radner W. Reading charts in ophthalmology. *Graefes Arch Clin Exp Ophthalmol*. 2017;255:1465-82.

10 Bach M, Kommerell G. Sehschärfebestimmung nach Europäischer Norm: wissenschaftliche Grundlagen und Möglichkeiten der automatischen Messung. *Klin Monbl Augenheilkd*. 1998;212:190-5.

11 Brussee T, van Nispen R, Klerkx E, Knol D, van Rens G. Comparison of reading performance tests concerning difficulty of sentences and paragraphs and their reliability. *Ophthalmic Physiol Opt* 2015;35:324–35.

12 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;254:177-84.

13 Radner W. Ophthalmic reading charts : Part 2: Current logarithmically scaled reading charts. *Ophthalmologe*. 2016;113:1029-1035.

Figure legend:

Figure 1:

RADNER Reading Chart. The print sizes of the sentences optotypes progress geometrically (logarithmically). In addition, charts with numbers and Landolt rings are included in the booklet.

Accepted Manuscript

