Evaluation of Reading Performance and Visual Acuity Tests

Mashaaer Abdulwahab Baashen

Doctor of Philosophy

Aston University

March 2020

Aston University

Evaluation of the Reading Performance and Visual Acuity Tests

Mashaaer Baashen

Doctor of Philosophy 2020

Thesis Summary

Improving reading ability is of high priority for most patients visiting optometric practice. Thus, it is expected that reading performance is one of the most important outcome measures for judging reading ability and the effectiveness of ocular interventions. Measuring near and distance visual acuity are simple and quick, but they cannot predict reading acuity, reading speed and critical print size which better reflect real-life reading performance. In contrast to distance visual acuity charts, there is not yet agreement on the best test to evaluate reading performance. There are many reading test charts available.

Reading test charts should be equally reliable. However, the work described in this thesis has shown that using different reading test charts, such as Radner, MNread, Colenbrander, Bailey-Lovie and IReST, resulted in different reading performance metrics. There are no studies that have undertaken a direct comparison between all these reading tests, hence in this thesis test-retest and inter-chart reliability of the reading test charts has been compared for pre-presbyopic, presbyopic and cataract subjects. Although the reading test charts presented in this thesis are considered as standardized tests in terms of the test item, the reliability results vary and can be classified as poor, acceptable and very good.

Using a reliable reading chart to evaluate the efficacy of the presbyopia treatment is a useful tool to investigate the reading performance rather than isolated letter near visual acuity. Automated measurement of the reading performance metrics by using a computer-based reading test could overcome the variation of the results between the practitioners but needs to be further calibrated. The findings of this thesis have a number of important implications for current and future ophthalmic practice.

Keywords: cataract, presbyopia, reading acuity, reading speed, reading test chart

Acknowledgement

I owe a special debt of gratitude to my supervisor, Dr. Shehzad Naroo, for his constant support, as well as giving me full confidence to effortlessly complete the work. He was the best supervisor anyone could wish to have. He was always ready to help, not just for the study, but also in making me feel that I have a second family here. He was an invaluable source of inspiration and knowledge. He convincingly guided and encouraged me to be a professional researcher and gave me the opportunity to expand my social network by attending and participating in some conferences.

I would like to deeply thank my associate supervisor, Prof. Frank Eperjesi, for giving me insightful comments and suggestions and warm encouragement during my PhD.

I would like to thank Dr. Fabrizio Zeri for his generous support and guiding me at the beginning of my PhD in how to conduct the proper methodology and statistical analysis. I appreciate the feedback offered by Prof. Stephen Anderson in the first and second year qualifying reports assessments and giving me the inspiration of some of the studies in this thesis.

I would like to thank the American Academy of Optometry (AAO 2018) for giving me an opportunity to present my research and share the knowledge with other expertise. The Saudi Royal Embassy and King Saud University are greatly appreciated for their support and funding. My thanks go to all the participants who took part in this work for their cooperation. Without their help and patience, this project could not have reached its goal.

Finally, I would like to offer my heartfelt thanks to my parents, sisters, brothers and friends for always being there for me, especially my brother, Dr. Mohammed Baashen. They kept me going and the work would not have been possible without their unlimited support.

Content

List of Figures

List of Tables

Chapter 1
Introduction and Literature Review

1 Introduction

Numerous visual assessment tests are available in ophthalmic clinical practice and practitioners may face a challenge in choosing the most appropriate test for a given assessment. The appropriate standardization of a given clinical test leads to valid interpretation and management, provides comparative information and useful primary results for a series of other clinical tests to determine the level of visual function (Cicchetti, 1994). Regardless of the structural specifications of the clinical test, statistical evaluation of that method in order to assess the reliability, repeatability and validity is important in order to discover issues related to the test's properties. Such standardization is fundamental for psychophysical tests, such as reading and visual acuity charts,

Since reading is a highly complex skill and involves the interaction of oculomotor, cognitive and sensorimotor perspectives (Latham and Whitaker, 1996), reading tests have become a useful investigative tool in clinical ophthalmic research, and in many other fields , such as neurology, psychology and psychiatry (Fulcher, 1997). Unfortunately, previous studies have not dealt with statistical evaluation of reading charts as much as they have with distance visual acuity charts.

In 1983 the International Council of Ophthalmology (ICO) specified the standard design principles of the reading chart in accordance to logarithmical progression (Colenbrander and Committee, 1988). However, a number of unstandardized reading charts have been developed, such as that printed for the purpose of advertisements of an ophthalmic company. The use of unstandardized reading charts has many disadvantages and may result in under or overestimating reading performance, the variety of the results and incomparability with other reading charts (Radner, 2017). Only a few reading charts have been developed according to ICO recommendations and these are as follows: Radner reading chart, Minnesota low vision (MNread) reading chart, Colenbrander reading chart and Bailey-Lovie word reading chart.

It is well-established that reading sentences or words is more difficult than reading isolated letters as in visual acuity tests (Bailey and Lovie-Kitchin, 2013). Accordingly, distance visual acuity measurements has been found to be a poor predictor of reading performance (Legge et al., 1992).

1.1 Clinical applications of the reading test chart

There have been attempts to standardize reading charts for the purpose of evaluating clinical outcomes. Reading performance achieved with monofocal IOLs, multifocal IOLs (Richter-Mueksch et al., 2002), refractive laser treatment of presbyopia (Baudu et al., 2013), and following LASIK/LASEK (Richter-Mueksch et al., 2005) has been evaluated with different types of standardized reading charts. In addition, evaluating the reading performance of patients with different types of cataract (Stifter et al., 2005b). It has been reported that a valid reading performance test has the ability to discriminate the visual impairment caused by age-related maculopathy and that caused by cataract (Stifter et al., 2004b). The Bailey-Lovie word reading chart was used to evaluate reading performance in cataract patients (Elliott et al., 2001) and the MNread reading chart was used to compare two types of accommodative IOLs (Brown et al., 2009), investigating the reading performance in albinism (Giacomelli et al., 2013), patients with retinitis pigmentosa (Virgili et al., 2004b) and diabetic macular oedema (Edington et al., 2017). The Radner reading chart was used to evaluate reading performance of patients affected by several eye conditions, including amblyopia (de Wit et al., 2012), age-related macular degeneration (Koch et al., 2012), diabetic retinopathy (Finger et al., 2009), uveitis (Kiss et al., 2006), macular hole surgery (Caramoy et al., 2011) and low vision patients (Burggraaff et al., 2012).

Results obtained with standardized reading charts allow more accurate comparison between research studies and comparison of clinical outcomes during follow up.

1.2 History of clinical reading charts

A number of ophthalmologists developed standards for visual acuity tests in the second half of 19^{th} century. In 1854, Eduard Jaeger von Jaxtthal developed the Jaeger chart, ''Schrift-Scalen'' (Runge, 2000) and his interest in the standardization of visual acuity measurement inspired the introduction of the Snellen distance visual acuity chart in 1862 using the principle of optotype construction (Snellen, 1862). After that, in 1867 and 1868, Green published the idea of logarithmic progression print size (Green, 1868). Thus, historic reading charts, such as Jaeger, which developed before the introduction of the design principle of the reading chart (Colenbrander and Committee, 1988) have lack of standardization. The historic reading charts were printed with limited print sizes due to the availability of rudimentary printing techniques (Radner, 2017). However, with modern printing techniques it is possible to achieve higher accuracy.

The Jaeger test chart consists of a series of sentence fragments on a decreasing in size (Fig 1.1). Although the Jaeger chart lost its original calibration, some of the reading test charts still use the Jaeger notation J1, J2, etc. (Jose and Atcherson, 1977). In addition, it has been found that different versions of English Jaeger charts are not comparable with each other and resulted in different reading outcomes (Colenbrander and Runge, 2007). Therefore, evaluation of the reading performance using a historic reading chart should be avoided for the purposes of research or in clinical examination.

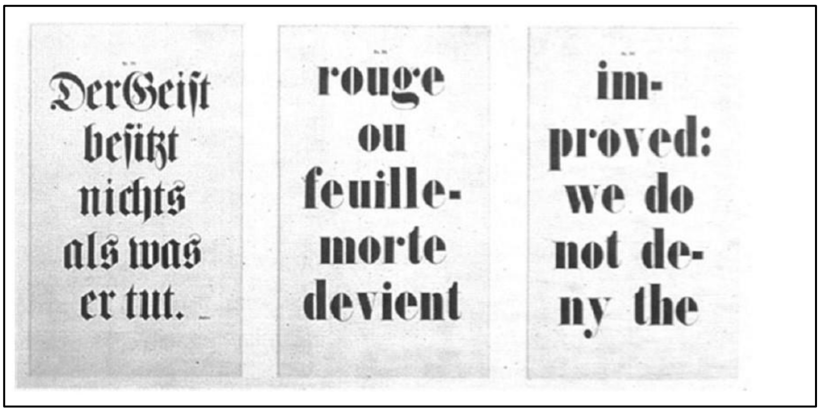


Figure 1.1: Original Jaeger reading charts in English, French and German (Runge, 2000).

In 1951, Law published the idea of N-notation to overcome the Jaeger chart limitations (Law, 1952). N-notation is based upon a point (pt) system and print sizes are not arranged logarithmically. The Sloan logarithmic progression reading chart was first described in 1963 (Sloan and Brown, 1963), followed by the Bailey-Lovie logarithmic progression word reading chart (Bailey and Lovie, 1980). A standard for reading acuity measurements was published in 1988 (Colenbrander and Committee, 1988). This includes: 1- print sizes should be arranged in a logarithmic progression; 2- test distance must be specified in all instances; 3- control of the variables that may influence the measurements such as inter-line spacing, letter spacing and typefaces; 4- the height of the lower-case letters subtends five minutes of arc; 5- the spaces between the words and the lines should be proportional to the letter size.

The following review explains some of the basic aspects for measuring reading performance and the background of the reading charts considered to be calibrated according to the ICO recommendations and outlines how far statistical evaluation has been performed on these reading charts.

1.3 Reading performance metrics

The reading performance metrics that can be analyzed and provide useful information about the visual function are reading acuity (logRAD), critical print size (CPS), reading speed, and logRAD/logMAR ratio. This ratio measures the reading acuity as a percentage of distance acuity. It has been found that logRAD/logMAR ratio is lower in people with age-related macular degeneration than normally sighted subjects (Radner, 2017) which affords this reading metric a useful diagnostic value.

1.3.1 Reading acuity

Reading acuity is expressed in terms of logRAD (= log Reading Acuity Determination), which is the reading equivalent of the distance logMAR. Reading acuity is mainly the smallest print size that can be read taking into account the number of the reading errors as a part of reading acuity (Radner et al., 1998). However, the methods for scoring the reading acuity vary between different reading charts. In the Radner reading chart, the number of reading errors is calculated based on the number of syllables of incorrectly read words (each syllable has a value of 0.005 logRAD), while, in the MNread, it is calculated on the basis of incorrectly read words (each word has a value of 0.01 logRAD). Other reading charts do not specify the value of each word and calculate the reading acuity based on the smallest print size only. In the Radner reading chart, incorrectly read words with more syllables (long word) has a higher impact on reading acuity than the short word (Radner et al., 1998). Therefore, using the standard word length to incorporate the number of the reading errors into reading acuity may reduce the variability of the results.

1.3.2 Reading speed

Reading speed in words per minute (wpm) is calculated from the number of the words in the sentence read correctly and the time needed to read that sentence (MacKeben et al., 2015, Alabdulkader and Leat, 2017, Han et al., 2017). However, other studies calculated the reading speed from the actual number of the words irrespective including those read incorrectly (Radner et al., 1998, Radner et al., 2002, Maaijwee et al., 2007). In addition, in some studies, the reading speed can be analyzed directly as the reading speed in words per minute (Radner et al., 2002, Maaijwee et al., 2008b, Radner and Diendorfer, 2014), while in other studies they performed a logarithmic conversion to log (words per minutes) (Virgili et al., 2004a, Subramanian and Pardhan, 2009, Alabdulkader and Leat, 2017). It has been found that logarithmic conversion reduces the variability of the results (Calabrese et al., 2016), and, thus, makes it difficult to judge or compare the reliability of the reading charts between different studies. Calabrese et al. (2016) has advised that reading speed should be reported in wpm rather than log (wpm) as it is more informative and understandable for practitioners.

For every study, maximum reading speed (MRS), different terminologies and scoring rules were used. The MRS is called optimum reading speed in some researches. The MRS was calculated using four different scoring rules. First, it was caculated as MRS of the sentences corresponding to or larger than CPS (Virgili et al., 2004a, Rice et al., 2005, Subramanian and Pardhan, 2006). Second, the MRS of the sentences corresponding to or larger than CPS, with the exception of the first two sentences (Maaijwee et al., 2008b). Third, average reading speed of the fastest three sentences read (Patel et al., 2011). Fourth, it was calculated as the speed of the fastest read sentence and called absolute reading speed (Burggraaff et al., 2010b). This disagreement on what scoring rule should be used makes direct comparison between studies complicated. It has been shown that the different scoring rule methods influence the reliability and the repeatability scores of the reading charts (Hazel and Elliott, 2002).

1.3.3 Critical print size

Critical print size measurements are based on a model of the graph plotted of reading speed versus print size. Reading speed of a person remains constant over a wide range of print sizes, at one point, when the print sizes become smaller, the reading speed decreases rapidly before reaching the acuity limit. This point is called CPS and is defined as the smallest print size that the patient can read with a MRS (Subramanian and Pardhan, 2009). However, for measuring CPS, the following six scoring rules were used in previous studies:

1. The print size at which the following smaller print sizes were read 1.96 SD slower than the MRS of the previous print sizes (Virgili et al., 2004a, Subramanian and Pardhan, 2006, Subramanian and Pardhan, 2009, Patel et al., 2011, Mataftsi et al., 2013)
2. Print sizes were read 5% slower than the mean reading speed of the previous print sizes (Virgili et al., 2004a, Patel et al., 2011)
3. CPS was the smallest print size that supported 90% of the MRS (Patel et al., 2011).
4. CPS was the smallest print size that supported 80% of the MRS (Patel et al., 2011).
5. CPS was calculated as the smallest print size that could be read with optimal or MRS (Stifter et al., 2004b, İdil et al., 2011).
6. The steepest decline slope on the graph of plotted reading time in wpm per sentence for each patient was used (Rice et al., 2005, Maaijwee et al., 2008b, Burggraaff et al., 2010b).

Patel et al. (2011) reported that MRS and CPS varied depending on the scoring rules applied to calculate them (Figure1.2). Also, it was found that CPS reliability will be affected when calculated through different scoring methods (Burgraaff et al., 2010).

Figure 1.2: Illustrative example of the scoring rules used to calculate CPS and MRS. (A) MRS (158 wpm) was calculated as the speed of the sentences in print size larger than CPS and CPS calculated as the smallest print sizes were read 1.96 times standard deviation slower than the average of the reading speed of the previous sentences. (B) MRS was calculated as the mean reading speed of the fastest three sentences (173 wpm) and the CPS scoring rule was as illustrated in the figure above. (C) MRS was calculated as the speed of the fastest sentence read (180 wpm) and the CPS was as illustrated in graph C. Reproduced by Patel et al. (2011).

1.4 Reading chart notations

For reading charts, different notations are currently in use, such as Snellen, logRAD, the N notation, M-Size and Decimal. Different notations are used worldwide depending on the practitioners' educational background and location.

M-Size. M notation was introduced by Sloan (1959); the letter height subtends five minutes of arc at a distance of 1 m but is also used for the reading distance of 40cm. One disadvantage of M notation is that the value of the notation is arranged in descending order (from 6.3M for the largest print size to 0.32M for the smallest print size), while the other reading notations, such as decimal, is arranged in ascending order.

N notation. This was first introduced by Law (1952) to overcome the limitations of the Jaeger chart. N notation is based on the point (pt) system and uses the Times New Roman typeface. Although the N notation is a familiar system because it represents everyday reading material typeface, different versions have different letter height and it does not correspond well to the other reading notations. Also, the N notation does not apply logarithmic progression and is not developed according to the ICO standards. Therefore, it could be reasonable to avoid using the N notation.

Snellen fraction. The Snellen fraction notation indicates the relation between the distance at which the letter subtends five minutes of arc and the test distance, in metres or feet. The

Snellen fraction of 20/20 (feet) or 6/6 (metre) could represent the acuity at distance and at near. Therefore, Bailey and Lovie-Kitchin (2013) have suggested using 0.4/0.4 Snellen fraction for a test distance of 40cm (0.4 metres). However, this fraction is not commonly used and using the Snellen fraction is more likely to be understood with provides clear information about the test distance.

Decimal notation. Decimal notation is a reciprocal proportion of the visual angle that is calculated from letter height and test distance. Whatever the test distances used, the decimal notation reveals a correct logarithmic progression.

LogRAD. LogRAD notation has been suggested to be used for the reading acuity equivalent of the logMAR to avoid confusion between reading and distance acuities (Richter-Mueksch et al., 2002). However, similarly to distance acuity, reading acuity can be expressed in logMAR units.

VAS scale. The Visual Acuity Score (VAS) scale was introduced by Bailey and Lovie (1980). The VAS scale is designed to simplify the rating of visual acuity deterioration as it is a scale of functioning. Each letter read on an ETDRS chart equals 1 point, so that each line read equals to 5 points. LogMAR (0.00) corresponds to 100 on VAS scale, 1.0 logMAR corresponds to 50 on VAS scale, and 2.0 logMAR corresponds to 0 on VAS scale.

1.5 Reading performance test charts

There are many types of reading test charts used in research , or in clinical practice, such as the unrelated words reading test – which was developed by Bailey and Lovie (1980) – continuous text (meaningful) reading tests, mixed contrast sensitivity tests, and computerized reading tests. In fact, the term (continuous text reading test) has a slightly confusing meaning. In this thesis in short sentences reading tests, this term is meant to refer to meaningful sentences. However, each sentence is independent to each other and do not make a full story as in a long paragraph reading test.

The following paragraphs provide an overview of the measurement properties and the characteristics of reading test charts that are available in the literature and calibrated according to ICO recommendations. Although other reading tests have been found, such as the Fonda-Anderson reading chart (Fonda and Anderson, 1988), the Sloan reading cards (Sloan and Brown, 1963) and the Zeiss reading test (Nguyen et al., 2009), no publications or relevant information have been found regarding their measurement properties.

1.5.1 The Bailey-Lovie word reading chart

Bailey and Lovie (1980) developed the Bailey-Lovie word reading chart with logarithmic print size progression and used sentences with unrelated words. The Bailey-Lovie reading chart was designed to determine the reading speed, reading acuity and CPS in one simultaneous examination. The same principle has been also applied to the MNread, Colenbrander and Radner reading charts. The reading notations labelled on Bailey-Lovie chart are M-units, N notation, LogMAR and VAR. The print sizes values are given for test distance of 25cm and use Times New Roman typeface (serif font). Therefore, a correction factor needs to be subtracted from the labelled logMAR for testing at longer distances. Bailey and Lovie viewed that measuring the reading performance with unrelated words led to more reliable results than continuous text reading chart. From their observation, they found that measuring the reading performance with unrelated words eliminates the influence of contextual cues, guessing, and individual variation of the reading behavior to affect the results. Thus, care was taken to eliminate any syntactic association between the words (Bailey and Lovie-Kitchin, 2013). Although the Bailey-Lovie charts are criticized in terms of the word length (more than 10 letters) and being difficult to read for poor readers, they are still widely used for determining the magnification power for low vision patients (Rubin, 2013). No studies were found in the literature that evaluate the Bailey-Lovie unrelated words reading chart characteristics in terms of validity, reliability and repeatability. Also, a direct comparison of the reading performance between the Bailey-Lovie reading chart and the available continuous text reading charts has not been investigated.

1.5.2 The Minnesota low vision (MNread) reading chart

The MNread reading test was first introduced in 1989 in the USA and was designed as a computer-based test (Legge et al., 1989). The sentences were presented on a computer screen with very large print sizes to low vision patients. Mansfield et al. (1993) then designed a printed MNread acuity card arranged in a logarithmic progression with the text difficulty being for third grade (age nine years). The current MNread chart is available in different languages (English, Italian, Greek, Spanish, Japanese, Portuguese and French) with a 3rd to 4th grade level of text difficulty (age 9–10 years) and with controlled position of the words and equal number of characters. The MNread test contains 60 characters, including the spaces, and applies the concept of “standard-length words” (Mansfield et al., 1993). The standard-length words are defined to have six characters in the English language, thus, each sentence in the MNread chart consists of ten standard-length words (three lines per sentence). It has been found that using the standard-length words helps to reduce the variation in scoring caused by different word lengths (Carver, 1976). The reading notations labelled on the MNread chart are

Snellen, M-units and logMAR notation for 40cm reading distance. The MNread chart has two editions to avoid learning effects during the test and uses Times New Roman typeface.

A number of studies have assessed the MNread internal consistency in different languages. For example, Mataftsi et al. (2013) developed and assessed the Greek version of the MNread acuity card and found a good internal consistency (Mataftsi et al., 2013). The same results were found in the Italian MNread (Vergil et al., 2004), Turkish (İdil et al., 2011) and Portuguese versions (Castro et al., 2005). Subramanian and Pardhan (2006; 2009) reported a good coefficient of repeatability of the MNread chart in visually impaired patients and in young adults; however, Patel et al. (2011) subsequently found weaker coefficient of repeatability in age-related macular degeneration patients. The reason for the differences between the studies was the variability in the study design and patient-related aspects.

1.5.3 The Radner reading chart

Radner et al. (1998) developed the Radner reading chart in Austria and adopted the logarithmic progression of the sentence optotypes to simultaneously determine reading acuity, reading speed and CPS. The Radner chart consists of 14 sentences and each sentence consists of 14 words. The sentences are displayed on three lines and comparable to each other in terms of number of characters, number of syllables per word, word length, positions of words and lexical difficulty with 3rd to 4th grade (9–10 years) (Richter-Mueksch et al., 2002). A san-serif Helvetica typeface is used for the Radner charts, which are available in English, German, Dutch, Hungarian, Spanish, Italian, French, Swedish, Turkish, Danish and Portuguese. To avoid learning effects during the test, , three different editions are available in the Radner chart. Reading notations (M-units, Snellen, decimal and logRAD) were labelled for 32cm and 40cm reading distances. The Radner reading chart was investigated in terms of test-retest and inter-chart reliability in normal sighted subjects and low vision patients. The results revealed good repeatability and reliability for measuring reading performance (Stifter et al., 2004a, Maaijwee et al., 2008b).

1.5.4 Colenbrander reading chart

The Colenbrander reading chart was developed by August Colenbrander in San Francisco in 1994. A variety of cards were developed by Colenbrander from 1994 to 2008, including low vision cards with the largest letter size being 20/500 at 40cm, small cards with 40cm cord, and low-contrast and mixed-contrast sensitivity cards. The Colenbrander high contrast continuous text print sizes use logarithmic progression and are used to measure reading acuity, reading speed and CPS and are available in different languages (English, Portuguese, Spanish, German, French, Danish, Dutch, Swedish, Italian, Finnish and Tagalo). The Colenbrander chart uses 14 pairs of sentences with equal length and spacing. For Snellen acuities from 20/230 to 20/200, one sentence is presented for print size, while, from 20/160 and smaller, two sentences are presented. it was created according to the criterion that there should be no more than 10 letters in one word, and it was not tested for text difficulty; however, later on, it was found to have a 4^{th} grade level of difficulty (Calossi et al., 2014). The Colenbrander chart uses Times New Roman typeface. Snellen, decimal and M-units reading notations were labelled on the Colenbrander reading chart for 40cm reading distance.

Although the inter-chart reliability, repeatability and reproducibility of the Colenbrander reading cards have not been investigated, a number of studies have developed new continuous text reading charts in different languages based on the Colenbrander chart criteria. Examples include the study by Abdulkader and Leat (2017) who developed a standardized Arabic reading chart with a layout similar to that of the Colenbrander chart.

1.5.5 International Reading Speed Test (IReST)

The International Reading Speed Tests (IReST) design is very different compared to the MNread, Radner, Colenbrander and Bailey-Lovie reading charts, it is available in long text paragraphs and uses one print size (1M or 10-point Times New Roman font). Patients with near visual acuity poorer than N8 (newsprint size) can read the paragraphs with magnifying aids. The advantage of the long text paragraphs is that they reflect the real-life reading performance in terms of reading fatigue, fluency and making mistakes (Stangler-Zuschrott, 1990). It has been reported that there are no differences in the reading speed between the short sentences and the longer text paragraphs (Radner et al., 2002). However, other studies have reported that reading speed measured with longer paragraphs (IReST) are more reliable than those with shorter sentences, as in the Radner and the MNread, because of the lower variance (Trauzettel-Klosinski et al., 2012).

The International Reading Speed Tests were developed specifically for measuring reading speed across languages to make the comparison in international studies; they are now used in the evaluation of reading performance studies if different languages are being compared in terms of age-related macular degeneration, dyslexia, magnifying aids and eye movement. The grade of the text difficulty ranges from 4th to 6th (age 10-12 years). The test is available in 18 languages, Chinese, Arabic, English, French, Finnish, German, Italian, Hebrew, Japanese, Polish, Dutch, Portuguese, Brazilian, Slovenian, Swedish, Spanish, Turkish and Russian (Hahn et al., 2006), and recently a Greece version has been developed by Gleni et al. (2019). The IReST test has a high translinguistic compatibility and scores well in 18 different languages and has scored high in terms of inter-language comparison studies (Hahn et al., 2006, Trauzettel-Klosinski et al., 2012).

1.5.6 Computerized reading test

Recent developments in computer-based technology provide a valuble opportunity for health practiononers and researchers to develop computer-based clinical tests which can combine multiple tests in one portable device (Berger, 2010). Reading performance test is one of the clinical tests that may benefit from the use of the new technology. The Salsburg Reading Desk (SRD) developed by Dexl et al. (2010) displays the text on a computer screen (Fig 1.2) and also has a voice detection to measure the reading time. The SDR can be adjusted to the reader’s preferred viewing distance. In addition, the reading performance test would benefit from tablet-based computer, such as Apple iPad, by increasing portability and standardizing the scoring method with automatic determination of the reading performance metrics. Examples of the iPad reading test are MNread iPad app (Calabrese et al., 2014), Radner reading app (Kingsnorth and Wolffsohn, 2015) and IUread reading test (Xu and Bradley, 2015). However, there is only one study available in the litrature for each iPad reading test and this was conducted by the developers. They reported a good reliability and repeatability of these tests for measuring the reading performance.

Figure 1.3: The Salzburg Reading Desk uses a computer-based technology to measure the reading performance metrics by presenting text in a random order. Reproduced by https://www.srdvision.com/

All the reading test charts mentioned above are designed to offer an objective method that examine out loud reading speed. There are other reading tests that are designed to evaluate silent reading over sustained period of time. However, it has been found that sustained silent reading tests do not accurately reflect out loud reading tests (Ramulu et al., 2013). Silent reading test is usually accompanied with comprehension questions corresponding to a reading passage in order to avoid skipping some reading material and to minimize the variation in attention given to the reading passage; thus, controlling this source of variability. One disadvantage of sustained silent reading test is the time consuming as the duration of the test may take up to 30 minutes, and this may not be feasible in research setting or in clinical practices where the time is restricted. Despite this, silent reading represents the large majority of our normal reading behavior.

1.6 Design aspects of the reading test charts studies

A number of studies have assessed the MNread internal consistency in different languages. For example, Matafsi et al. (2013) developed and assessed the Greek version of MNread acuity card and found a good internal consistency in the chart (Matafsi et al., 2013). The same results were found in the Italian MNread version (Vergil et al., 2004), in Turkish (Edil et al., 2011) and Portuguese version (Castro et al., 2005). The reliability of the current MNread in English has not yet been assessed. The internal consistency of the Radner chart has been tested in German (Radner et al., 1998), Spanish (Alio et al., 2008), Dutch (Maaijwee et al., 2007) and English (Radner and Diendorfer, 2014). The internal consistency of IReST was also evaluated in different languages among young normally sighted subjects (Trauzettel-Klosinski and Dietiz, 2012) as well as presbyopic subjects (Hahn et al., 2006).

However, different word length resulted in different languages (Delattre, 1965) and this should be taken into account, especially given that the reading performance results are based on the model of reading speed against print size. Therefore, direct comparison between reading speeds across reading test chart studies in different languages is complicated. Longer words slow reading speed as more reading material is covered in a long words sentence compared to a short words sentence (Brussee et al., 2015).

Reading performance measurements are influenced by nonvisual factors, such as age, reading habits, practice effect, educational level, motivation, attention, fatigue, visual acuity and disease-related factors, such as stage, duration and type of the eye disease (Finger et al., 2009, Sunness and Applegate, 2005, Legge et al., 1992). In addition, other experimental designs, such as the illumination level and the reading distances, play an important role in the variability of the reading test results. It has been reported that a small change in the reading distance may lead to different reading performance metrics results (Patel et al., 2008). There is a strong relationship between the illumination level and the subject’s ability to distinguish the letters during the measurement of near visual acuity and this effect increases with age as a result of age-related changes of the ocular media, leading to decrease in the retinal illuminance (Haegerstrom-Portnoy et al., 1997). It has been shown that normally sighted subjects experience no or minimal improvement in the reading performance with illuminations levels, whilst the low vision subjects experience a noticeable improvement (Cornelissen et al., 1994). Munch et al. (2016) reported that education level had an influence on the reading performance results. They found that university students read faster than factory workers and made a lower number of reading errors.

Other important factors that are important to consider when measuring reading performance are the visual crowding and the eye movement during reading. In human eye, the retinal fovea is represented by the central two degree of the visual field with the highest cone density. The fovea is responsible for the tasks that require resolving fine visual details. During reading, the eye continuously makes a series of saccades (bringing the object of interest to the fovea when the eye moves from one location to another) and fixations (when the eye remains fixed and stable) (Rayner, 1998). Fluent readers make regression (back to previously read word or text) about 10% to 15% of the time (Starr and Rayner, 2001). Both fixation and saccades length fluctuate from word to word and are affected by syntactic, word frequency and number of lexical (Balota et al., 1985).The rest of visual information falls on the peripheral retina. One of the most disruptive deficits in the peripheral vision are those correlated to visual crowding. Visual crowding is the inability to identify the letter in clutter in the periphery (Nandy and Tjan, 2012). Therefore, visual crowding is more sequential for people who rely on their peripheral visual field for reading, as they do not have a functioning fovea. Understanding the eye movement and reading process in general will help to a better measurement of reading performance.

Table 1.1 illustrates data extracted from available studies that assessed the reading test charts. Some of the design aspects of some studies are missing or not clear, which makes the comparison between the outcomes difficult. The design aspects of some studies will be illustrated in the following paragraphs.

The Radner and MNread have been investigated in normal and low vision subjects while the IReST has been tested in normally sighted subjects. In previous studies, some specifications and participant characteristics, such as state of the eye disease or educational level, were not mentioned. However, a number of studies have described their methods very clearly, such as the study conducted by Maaijwee et al. (2008) to analyze the inter-chart and test-retest reliability of the Dutch version of Radner reading charts in an older group of subjects with macular diseases. In the design aspects of their study, they mentioned the eye disease state, the recording method, controlling the reading distance and lighting level as well as that described in studies of Radner et al. (2002) and Subramanian et al. (2006).

Legge et al. (1989) evaluated the old computer-based MNread test for normal and low vision subjects. In their method, they did not mention the age of the subjects, the educational level, the method for controlling the reading distance, the distance or near visual acuity, nor the recording methods (Legge et al., 1989). The study by Mansfield et al. (1993) did not mention their method or the same aspects as Legge et al. (1989) in addition to an unspecified inter-session period (Mansfield et al., 1993). Although Ahn et al. (1995) described their study design well in terms of the subjects' age, visual acuity, state of the eye disease, lighting level and recording methods, the way of controlling the reading distance was not clear. All these studies measured reading speed, which is strongly affected by age, distance, near visual acuity, and illumination level (Haegerstrom-Portnoy et al., 1997).

In the context of the design aspects of the scientific studies, Rice et al. (2005) evaluated a new abbreviated version of the MNread reading speed test in children with normal vision and in children with a variety of ocular conditions; however, they did not describe the state of the eye disease that the children had. Another study investigated the repeatability of the recent MNread charts with low vision subjects (Subramanian and Pardhan, 2009). However, they let the subjects use their preferred viewing distances, which may make comparisons with their results complicated.

It is worth noting the difference in the MNread repeatability results of Subramanian and Pardhan (2009) who scored better in repeatability compared to Patel et al. (2011), and both studied the repeatability of the reading performance metrics in the English language in low-vision subjects. The reason might be that Subramanian and Pardhan (2009) made use of the audio recording method to calculate the reading time and speed, which may lead to higher repeatability results, whilst Patel et al. (2011) made use of a direct notation using the stopwatch in the same clinical setting. The influence of the recording method on the reading performance metrics results will be investigated in this thesis.

Testing the reading test chart according to the scientific standards is an important consideration that help to choose the reliable reading test chart. As the reading performance metrics measurement are influenced by patient and measurement related factors, the studies should provide sufficient information to understand how the study was conducted and designed and how the results were analyzed.

1.7 Methods for scoring the reading performance metrics and their reliability results

In the current thesis, the term repeatability or test-retest reliability refers to variability in repeated measurement by one examiner when controlling for all other factors that may affect the results. Reliability or inter-chart reliability refers to the internal consistency of the item on a reading chart.

To date, there has been a little or no agreement on the scoring rule methods for measuring reading performance metrics and the statistical approaches for determining repeatability, reliability and the agreement of clinical reading test charts. Although, it has been reported that the use different scoring methods influences the reliability results (Burggraaff et al., 2010b, Patel et al., 2011, Hazel and Elliott, 2002), there is no agreement on the method of choice to calculate the reading performance metrics. Different methods were used to calculate reading acuity, MRS and CPS, which makes the comparison between studies difficult. In addition, there are many different statistical methods used in the studies to evaluate the reliability and the repeatability of the reading test charts. Investigating the reliability of the reading test chart is an important step because, when the reliability is high, a decrease of the reading speed and an increase of the reading errors is more likely to be related to disease rather than to the chart design or other characteristics (Radner et al., 2002). A high Cronbach's alpha indicates a good reliability or internal consistency of the reading chart (Tavakol and Dennick, 2011).

In the review conducted by McAlinden et al. (2011), they discussed the statistical methods for analyzing the agreement (comparison of the repeated measurements or between clinical tests) for ocular instruments. The Pearson correlation coefficient is most commonly applied in agreement studies. However, they argue that the correlation coefficient incorrectly assesses the agreement between the results because it is used to assess the linear relationship between the two variables and cannot assess the relationship beyond its linearity. In addition, the correlation coefficient is very dependent on the variability between subjects. Therefore, they concluded that the use of correlation coefficient in agreement studies is questionable as a strong correlation does not always indicate a good agreement. According to McAlinden et al. (2011), the method of choice to assess the agreement is the Bland-Altman agreement analysis (Bland and Altman, 1986) and also it is the most popular method to evaluate the agreement (Zaki et al., 2012). The limit of agreement at which it is acceptable to use between two methods or clinical tests interchangeably is a clinical decision that can be considered.

Based on the available studies that assessed the reading charts' inter-chart and test-retest reliability, the content and the quality of the method and study design used varied, making the comparison between the studies complicated. Therefore, it is still not clear whether one reading test had better repeatability and reliability than the other. One aim of this thesis was to investigate the reliability and the repeatability of the available standardized reading test charts (Radner, MNread, Colenbrander, Bailey-Lovie word chart and IReST), in addition to the computerized reading test, in young normally sighted subjects, presbyopic subjects and in patients affected by cataract. Taking into account control over study design, statistical approaches and scoring rule method, the findings should make an important contribution to optimize the reliability of the reading test charts to obtain comparable outcomes of the reading performance in patients suffering from various diseases. Also, the use of reliable reading charts facilitates research concerning the functional vision in many fields of eye care.

Table 1.1: The design aspects of reading test charts studies

Authors and publication year	Reading test	Language	Age of subjects (mean)	Subject criteria	Educational level	Distance visual acuity (logMAR)	Eye disease	Eye disease stage	Inter-chart reliability	Test-retest reliability	Lightning (cd m²)	Reading distance controlled
Radner and Diendorfer (2014)	Radner	English	30	Normally sighted	UV, non-academic	≤0	NA	NA	+	0	80-90	Yes
Burggraaff et al. (2010a)	Radner	Dutch	81	Low vision	8.5 years of education	≥ 0.3	Described	?	+	0	80-90	?
Maaijwee et al. (2008b)	Radner	Dutch	71	Low vision	?	≥ 0.4	Described	?	+	0	80-90	Yes
Alió et al. (2008)	Radner	Spanish	31	Normally sighted	UV	≤0	NA	NA	+	0	80-90	?
Maaijwee et al. (2007)	Radner	Dutch	?	Normally sighted	UV	≤0	NA	NA	+	0	80-90	?
Stifter et al. (2004a)	Radner	German	66	Low vision	?	≥ 0.4	Described	Described	+	+	80-90	?
Radner et al. (2002)	Radner	German	24	Normally sighted	UV	≤0	NA	NA	+	0	80-90	Yes
Radner et al. (1998)	Radner	German	24	Normally sighted	UV	≤0	NA	NA	+	0	90-100	?
Mataftsi et al. (2013)	MNread	Greek	32	Normally sighted	?	≤0	NA	NA	+	+	?	?
Patel et al. (2011)	MNread	English	78	Low vision	?	≥ 0.2	Described	?	0	+	120	?
İdil et al. (2011)	MNread	Turkish	?	Normally sighted	PE, HS, HE	≤0	NA	NA	+	0	80	?
Subramanian and Pardhan (2009)	MNread	English	73	Low vision	?	≥ 0.12	Described	Described	0	+	102	?
Subramanian and Pardhan (2006)	MNread	English	23	Normally sighted	UV	≤0	NA	NA	0	+	≥ 80	Yes
Rice et al. (2005)	MNread	English	11	Ocular condition	?	≥ 0.1	?	?	+	0	?	?

Table 1.1: (continued)

Authors and publication year	Reading test	Language	Age of subjects (mean)	Subject criteria	Educational level	Distance visual acuity (logMAR)	Eye disease	Eye disease state	Inter-chart reliability	Test-retest reliability	Lightning (cd m²)	Reading distance controlled
Virgili et al. (2004a)	MNread	Italian	11	Normally sighted	Grade 3-8	≤0.3	NA	NA	+	+	?	?
Ahn et al. (1995)	Old MNread	English	45	Low vision	?	≥ 0.12	Described	Described	+	0	?	?
Mansfield et al. (1993)	Old MNread	English	?	Normally sighted	?	?	NA	NA	+	0	?	?
Legge et al. (1989)	Old MNread	English	?	Low vision	?	?	Described	Described	+	0	100	?
Trauzettel-Klosinski et al. (2012)	IReST	13 languages	18-35	Normally sighted	?	?	NA	NA	+	0	?	?
Hahn et al. (2006)	IReST	4 languages	18-85	Normally sighted	?	≤0.2	NA	NA	+	0	?	?

NA, not applicable; ?, not determined; +, studied; 0, not studied; UV, University students; PE, primary education; HE, higher education; HS, high school

Chapter 2

The introduction highlights key areas of the routine eye examination that require further investigation. As a first step it would be useful to know why patients attend for a routine eye examination. In this chapter this will be explored.

Study 1: Why do we need to address the most common reasons for visiting optometric practices?

2 Introduction

Valid measures of the prevalence of the most common reasons for visiting optometric practices are the basis to improve the quality of care and services. The improvement of the practice and services offered comes only when they are measured (Peabody et al., 2004).

The chief complaint 'is a patient's self-reported primary reason for presenting for medical care' (Mowafi et al., 2013) and is also referred to as 'reason for visit' (RFV). Analysis of primary medical care should take into account the aspect of the chief complaint. Collecting a data on patients' chief complaints would help as primary data to analyse, quantify and plan for medical care system development (Begier et al., 2003).

A patient's chief complaint can be considered as a good resource to identify the need for service improvement (Shah et al., 2008). Unfortunately, there is a lack of research that takes the prevalence of the chief complaint into account to improve the standards of the services delivered to the patient. Patient complaints are usually used to address the patient's issue individually (Hsieh et al., 2005). However, the communication between the practitioners about the prevalence of the most common chief complaints and how to handle them can be used as a good source of learning and research that could help to improve the quality of care.

2.1 Categorization of the visual complaints among the optometric practices

Some practitioners would ask the patient about his or her chief complaints before questioning on his or her ocular history and some of them after. Some of the common chief complaints in optometric practice are discussed below.

2.1.1 Complaint of blurred vision

Blurred vision is one of the common symptoms related to the eye and it has a wide variety of causes. Sufficient history information and essential eye examinations can lead to accurate diagnosis of the cause of the blurred vision. The causes of blurred vision can be divided into two main categories, non-pathological and pathological causes (Shingleton and O'Donoghue, 2000).

2.1.1.1 Blurred vision non-pathological causes

2.1.1.1.1 Refractive error

One of the most common causes of blurred vision is refractive errors, which include hyperopia (farsightedness), myopia (nearsightedness), presbyopia (age-related decrease in accommodative response) and astigmatism.

Refractive errors have been confirmed globally as one of the most common causes of visual impairment (Resnikoff et al., 2008). Low vision resulted from uncorrected refractive errors can lead to very bad consequences for individuals in a population. For example, reduce quality of life, lost education, lost economic gain and employment opportunities. Some reasons why refractive errors remain uncorrected include lack of awareness about the importance of routine examination at public health level and family and personal level. Another reason is insufficient refractive services available (Resnikoff et al., 2008).

Usually, in many cases, the practitioner’s diagnoses of the cause of the blurred vision must await the refractive examination results. After that, different possibilities could be considered as the cause of the blurred vision complaint (Grosvenor and Grosvenor, 2007).

2.1.1.1.2 Functional vision loss

Another cause of non-pathological blurred vision is the blurred vision that results from psychological or perceptual malfunction and this is called functional vision loss (Shingleton and O'Donoghue, 2000). Griffiths and Eddyshaw (2004) suggested to use the term “*medically unexplained visual loss”* as a diagnostic label of functional vision loss. Constricted visual field and reduced visual acuity are the most common symptoms of functional vision loss. These symptoms are real, but usually transient. Treatment of functional vision loss needs an evaluation of the underlying psychopathology (Pula, 2012).

2.1.1.2 Blurred vision pathological causes

Blurred vision that results from pathological causes is less common than blurred vision that results from non-pathological causes, but is more urgent. Blurred vision complaint that is confirmed by decreased visual acuity and not caused by change in refractive errors usually needs additional testing and diagnostic procedure (Grosvenor and Grosvenor, 2007).

Symptoms that help in the diagnoses of varied causes of pathological blurred vision include the presence of the pain, bilateral or unilateral, and the onset of the blurred vision, whether suddenly or gradual. The structures of the eye that are sensitive to the pain are the ciliary body, the iris and the interior ocular surface as well as the optic nerve and periorbital areas. Therefore, the inflammation of these areas can cause blurred vision and pain. The pathological diseases of the vitreous and retinal can cause blurred vision, but are relatively insensitive to the pain (Shingleton and O'Donoghue, 2000).

2.1.2 Complaint of headaches

Headache in general and migraine in specific is one of the common reasons for visiting optometric practices. Analysis of the data collected from the patients for more than 15 years in a large sample of UK general practices found that 9% of the patients presented to general practices with a complaint of headache every year (Latinovic et al., 2006). In a study by Gutteridge and Cole (2000) conducted in 1000 consecutively presenting patients of optometric practice, the prevalence of migraine complaint was 11% of male and 23% of female.

Headache visual symptoms should be carefully diagnosed and differentiated from other ocular or neurological diseases. The optometrists need to be aware about the different types of migraine presentation. In many cases the optometrist may not find the reason for a headache and onward referral to a medical practitioner may be warranted.

2.1.3 Complaint of asthenopia

The definition of asthenopia is a visual sensation of eye strain or eye weakness and fatigue. Eye strain is a synonym of asthenopia, and it is a more common eye complaint in adults (Nakazawa et al., 2002) and children (Vilela et al., 2015), particularly among people who have jobs that need extended near vision, for example, computer users, accountants and secretaries (Nakazawa et al., 2002). Asthenopia can result from many different causes, such as convergence and accommodative anomalies, binocular vision anomalies, and uncorrected refractive errors including presbyopia and insufficient lightning. Asthenopia can be revealed through different eye symptoms such as double vision, watery eye, headache, blurred vision, headache, itching and dry eye (Neugebauer et al., 1992).

Because asthenopia results from different causes, the treatment involves the treatment of the work environment, such as the lighting and prolonged near vision work, as well as the treatment of the visual condition (Aarås et al., 1998).

Vilela et al. (2015) estimated the prevalence of asthenopia in children up to age 18 years through a systemic review and found that 19.7% of children complained from asthenopia while the majority of the frequency was found in children aged over 7 years. Previous studies in adults found that the frequency of asthenopia was more in those who used a video display unit (VDU) for a long period (Neugebauer et al., 1992, Bergqvist and Knave, 1994). Therefore, it is worth to note that children nowadays are heavy users of videogames, computers and smartphones, which may increase asthenopia prevalence. Asthenopia can cause additional consequences in school performance and learning.

2.1.4 Routine eye examinations

The need for regular eye examination is the key element to improving eye care. Many eye diseases, such as glaucoma and diabetic retinopathy, can progress and exist without the patient being aware of the problem until the vision is deteriorated or lost (Wensor et al., 1998). Recommendation for a routine eye examination usually targets the people with eye diseases and who notice a change in vision. However, Taylor et al. (2004) found evidence that people over 40 years old who had normal baseline eye examination developed a symptomatic vision loss over a five-year period. In addition, a modest reduction in visual acuity (<0.3 logMAR) may cause visual impairment and this could impact the quality of life and independent ageing (Weih et al., 2002). Therefore, the public health message about the importance of routine eye examination should target people with eye diseases as well as asymptomatic people.

Routine eye examination includes patients visiting the optometric practice for spectacles change or contact lens aftercare. Frequency of the routine eye examinations depends on the patient age and condition. The Collage of Optometrists in the UK recommended a two-year interval for people aged 16 years and older and a one-year interval for children up to 16 years old. According to the condition, the recommended minimum re-examination interval for diabetic people is one year and six months for children with binocular vision anomaly (The Collage of Optometrists, 2019). In general, more frequent routine eye examination is recommended for older adults and children than young adults (Irving et al., 2016).

Routine eye examination plays an important role to reduce the economic cost of vision loss by screening for symptomatic and asymptomatic eye diseases (Wittenborn et al., 2013).

2.1.5 Complaint of reading difficulty

Reading is an essential part of many daily activities and an important skill. Reading difficulty is the primary complaint in the elderly and individuals with eye disease (Mangione et al., 1998). It has been found that 60% of patients referred to low vision services complained about reading difficulty (Rubin, 2013).

Evaluation of the reading performance is an essential and important clinical test to maintain a good quality of life, especially for ageing population. As it has been found that reading difficulty can reduce and affect the quality of life (Alio et al., 2011).

The ageing and growth of the world's population is leading to an increase in the number of people with presbyopia who have complained about blurred near vision. It has been globally estimated that, of the 7.33 billion people alive in 2015, presbyopia affected 1094.7 million of people older than 35 (Bourne et al., 2017). Reading difficulty complaint may be accompanied with other visual complaints, such as eye strain and headache (Patel and West, 2007).

The need for adequate near vision is not only for reading and writing, many other tasks in daily lives also require good near vision. such as cooking food, sorting rice or grain, threading needles, weeding and dressing children. Thus, it demonstrates the impact of blurred near vision on quality of life in the developing countries (Patel et al., 2006).

The presence of reading difficulty complaint may associate with different variables. It has been reported that level of education, age, gender and social functioning are significantly associated with reporting reading difficulty complaint. However, reduced distance visual acuity was not associated with near vision task difficulty (Patel et al., 2006).

A valuable predictor for measuring the visual ability to perform daily tasks (McClure et al., 2000) and assessing the vision-related quality of life is a reading performance test (Hazel et al., 2000).

Therefore, if the information collected at the beginning of an eye examination and the reason of visit the optometric practices are investigated properly this may allow the optometrists to use the eye examination adapted to the patient's needs and in avoiding any irrelevant tests (Shah et al., 2008). In agreement with Anderson et al. (2001), such results and analysis of the most common chief complaints should be shared with healthcare professionals to facilitate the service improvement and development. The goal of this chapter is to identify the frequency and the type of the most common reasons for patients visiting an optometric practice in the UK and Australia. The secondary aim is to determine the relevance of age, gender, refractive errors and occupation to the presence of the visual complaints.

2.2 Methods

This study presents an audit of data collected at two optometric practices; a single practice in Adelaide, Australia; and a single practice in Nottingham, United Kingdom (UK). The two locations were both in suburban areas, not in the city centre, and were small group practices. The style of the practices was very similar, and both had high street locations in busy shopping areas. The data was collected in Adelaide in March/ April (autumn time in Australia); and in the same calendar year the data was collected in Nottingham but in September/October (autumn time in the UK). Adelaide has a larger population than Nottingham and is a bigger city, but the suburbs where the data were collected are of similar size and demographics (Highgate area of Adelaide and West Bridgford area of Nottingham).

The data were collected from consecutively presenting patients by means of an audit of the patient records. Records were selected from patients who had appointments at each practice. Importantly, data were collated from consecutively presenting patients so that no operator bias was involved in selection. The information recorded was gender, age, date of last eye examination, occupation, reasons of visit and refractive errors state. Reason for the visit of each participant were under the following complaints (routine eye examination, reading difficulty, distance vision blur, change in spectacles prescription, family history of the eye disease, pathology, headache and contact lenses). The initial visual complaint was considered as the reason for the visit.

The number of participants was 130 from Australia with mean age (mean± SD) 44.8± 20.4 years, 43% male and 57% female, and 110 participants from the UK with mean age (mean± SD) 40.5± 13.7 years, 43% male and 57% female. Data were entered and analysed using Microsoft Excel statistical software and Statistical Package for Social Sciences (SPSS) version 25. Chi-square analysis was used to compare the results across the two countries. Logistic regression was used to assess the relevance of the age, gender, refractive errors and occupation to the likelihood of presence of visual complaints.

2.3 Results

The results in Figure 2.1 indicate that the most common reasons for visiting an optometric practice in Australia were change in spectacles prescription (24%) and near vision difficulty (23%) followed by distance vision blur (19%). In the UK, the most frequent reason for visit was routine eye examination (35%), and near vision difficulty (23%) and change in spectacles prescription (19%) were the next most common complaints. Fewer than 10% of patients in Australia and the UK visited the optometric clinic with complaints that included pathology (Australia 5%; UK 2%), family history of eye disease (Australia 2%; UK 3%), headache

(Australia 9%; UK 10%) and contact lens prescription (Australia 2%; UK 5%). Overall, the complaint shared equally frequently and significantly in both countries was reading difficulty. Australian patients are more likely to have distance vision complaints compared to UK patients (19% vs. 2%; p = 0.02). The results showed that the awareness of UK patients about the importance of routine eye examinations was significantly higher than in Australian patients (35% vs. 17%; p = 0.03). The prevalence of change in spectacle prescription complaint was not statistically significant between Australia and the UK (24% vs. 19%; $p > 0.05$).

In these consecutively presenting patients, no statistically significant differences were found between the distribution of categorical age groups in Australia and the UK. The frequency of older age patients above 45 years was Australia 52% and UK 42%, for the younger patients whose ages were between 17 and 44 years old it was Australia 40% and UK 32%, and for children up to 16 years old, Australia 8% and UK 10%.

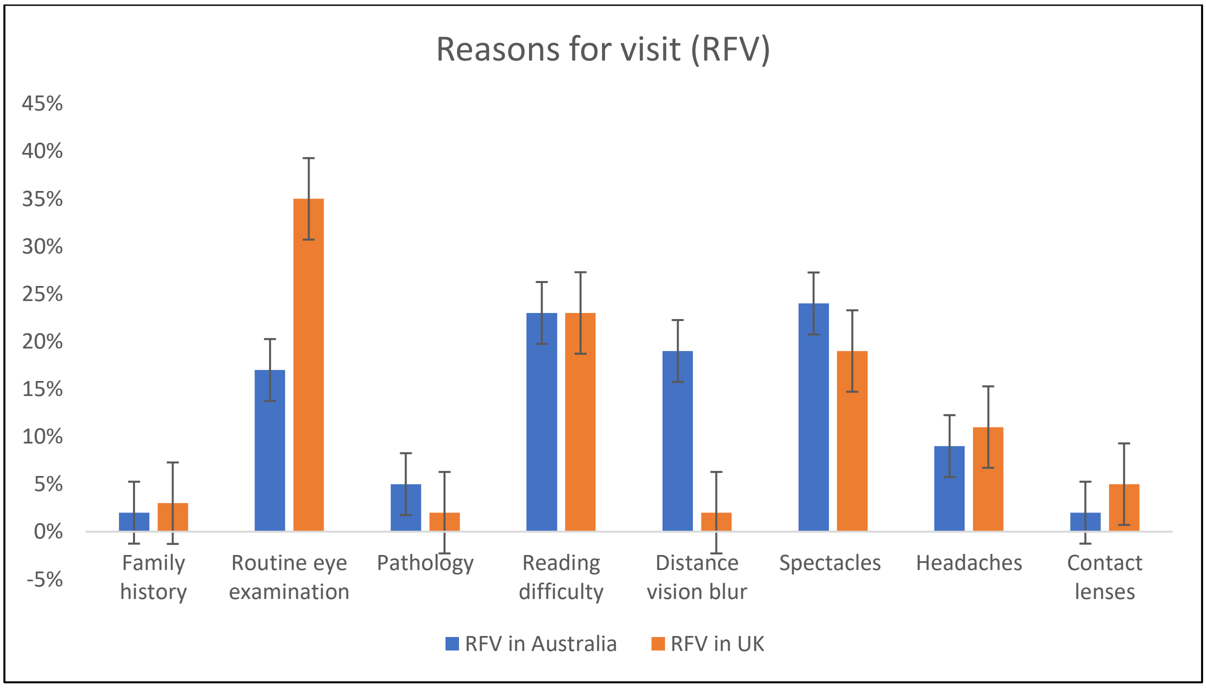


Figure 2.1: Chief complaints of patients seeking an optometric practice in Australia and UK

Although the older age (45-86 years old) was associated with greater frequency of the reading complaints, younger age participants (17- 44 years old) showed a higher percentage in presenting with near vision difficulties in Australia and the UK at 33% and 12%, respectively (Figure 2.2). In logistic regression analysis, no significant association was found between the age and the presence of reading complaints in Australia (odds ratio [OR], 1; 95% CI 0.99-1.03) and UK (OR, 1; 95% CI 0.98 -1.03). In Australia, older age was significantly associated with greater odds of visiting the optometric practice for routine eye examinations (OR, 1.109; 95% CI 1.013- 1.213), while, in the UK, no significant relevance was found between age and likelihood for visiting the clinic for routine eye examinations (OR, 1.040, 95% CI 0.986- 1.097). In both countries, older age was significantly associated with presence of eye pathology complaint, UK (OR, 1.396; 95% CI 1.031- 1.891), Australia (odds ratio [OR], 1.107; 95% 1.006 – 1.219). The frequency of change in spectacles prescription complaint increased with increasing age in Australia (OR, 1.090; 95% CI 0.998- 1.190) and the UK (OR, 1.057; 95% CI 0.998- 1.119). No significant association was found between gender, refractive errors or occupation and the presence of all visual complaints in Australia and the UK.

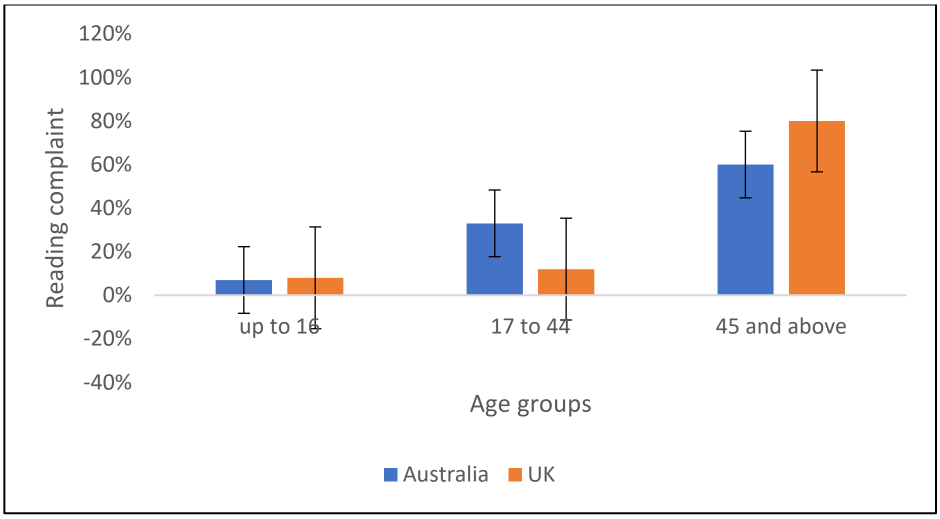


Figure 2.2: Age group complaint about near difficulty in Australia and the UK

The data obtained from the date of last eye examination for the patients revealed that the majority of the patients had an eye examination every two years (30% in Australia, 27% UK), three years (22% Australia, 17% UK) and one year (15% Australia, 13% UK).

For the Australian patients, the mean value of the spherical equivalent was (0.22 ± 2.50) and the frequency of myopia was 23%, hyperopia 27% and presbyopia 23%. The mean value of the spherical equivalent in the UK was (0.55 ± 2.12) and the frequency of the refractive errors myopia, hyperopia and presbyopia was 16%, 19% and 32%, respectively.

Figure 2.3 shows the frequency of patients' occupation in the two countries. The most interesting aspect of this graph is that the number of retired people visiting the optometric practice in Australia was significantly higher than those in the UK (30% vs. 7%; p = 0.0001). Also, the number of students in Australia (14%) was higher than those in the UK (3%). However, this difference was not statistically significant, p= 0.05.

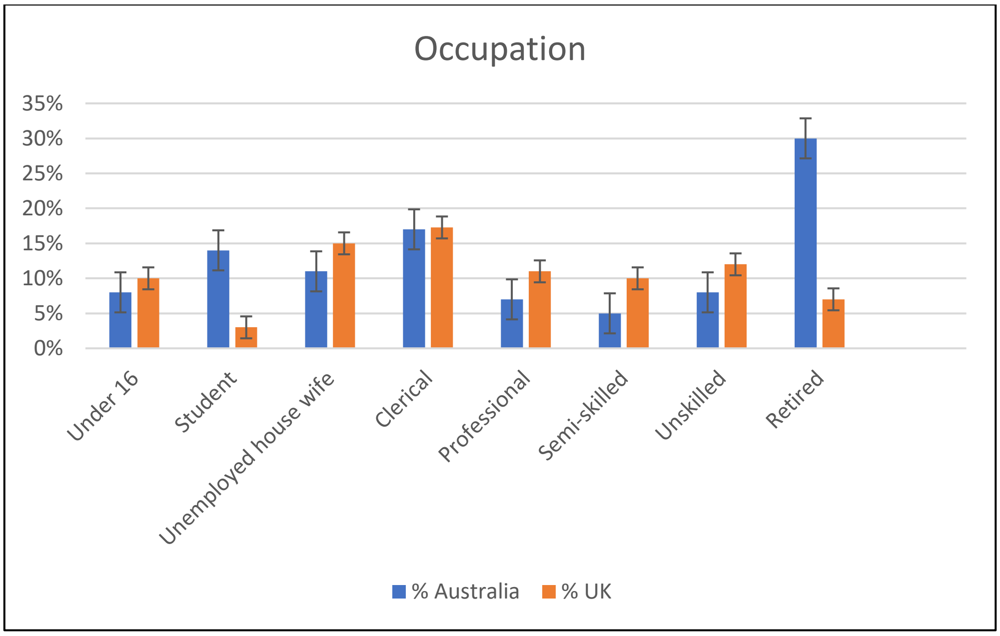


Figure 2.3: Comparison between the occupations of patients visiting an optometric practice in Australia and the UK.

2.4 Discussion

This study established the prevalence of different visual concerns in patients seeking optometric practices. Recognition of the most common reasons for visiting an optometric practice will provide clinicians and researchers direction in planning and development of the examination strategies needed to address the growing demand of vision problems. In this study, Australia and the UK were selected according to the progressive and advanced eye health services in these two countries (Hashemi et al., 2018).

This study was the first to assess the frequency of the main complaints of two groups of patients in seeking optometric practices in Australia and the UK. This study found that the frequency of the reading difficulty complaint was seen in both groups with an equally high percentage (23%). One possible explanation of the reading complaint prevalence is the term "reading", which covers a wide range of activities in daily life and with different degrees of difficulty, such as sustained continuous reading of magazines, books or a computer versus spot reading of bills, medication bottles or food labels. Furthermore, advanced technology, such as Smartphones, tablets and e-readers, have increased the number of formats in which reading takes place. In the developed societies, reading is fundamental skill to access information anywhere.

Understanding the importance of the reading tasks for patients is essential for evaluating and standardising the measures of the reading tests to address the whole spectrum of the reading performance. The results provide further support for the argument that reading tests should be standardised in all optometric clinics and fields, as the reading difficulty complaint is not associated only with older age group as a result of presbyopia and other aging ocular diseases, but also a percentage of children and students complained about reading difficulty. The presence of refractive errors or uncorrected refractive errors in this younger age group could be considered as a cause of reading difficulty. A population based study conducted to document the refractive error in a population of children in Australia found that the burden of refractive visual impairment was 10.4% (Robaei et al., 2006), and this was comparable with some East Asian countries where the prevalence of myopia is high, especially at school entry (Zhao et al., 2000). McCullough et al. (2016) found that the proportion of myopes among the children and teenagers in the UK has doubled in the last 50 years and the hyperopia above +3.50D tend to be stable and persistent. In spite of this growing demand of vision problems, the literature review suggests the lack of standard methods and reading test chart for measuring the reading performance in clinical practice within optometry.

Although this study was conducted routinely presenting patients for eye examination appointments, rather than for a specialist clinic, the results are in agreement with Brown et al. (2014) who found reading difficulty to be the most common reason for visiting low vision services in the United States. With respect to reading difficulty complaint, it is worth noting that the most common vision assessment questionnaires include one or more questions regarding reading difficulty.

In multivariate regression analysis, gender was not associated with likelihood of reading difficulty complaint. This outcome is similar to previous studies (Hazel et al., 2000, Brown et al., 2014).

Optometrists often advise patients that they must present for routine eye examination every two years. Some patients will follow this advice, whereas others may wait until they have a problem with their eyes or vision. Interestingly, the incidence of UK patients (35%) seeking an optometric clinic for routine eye examination was more frequent than Australian patients (17%). The awareness of UK patients about the importance of routine eye examinations was significantly higher than in Australian patients. It has been found that the percentage of five-year incidence of vision loss and complication in patients with a normal baseline examination is higher in people who have not had routine eye examination compared to people with frequent eye examination (Taylor et al., 2004). Even a modest visual reduction and uncorrected refractive errors may cause half of the presenting visual impairments (Dandona et al., 2002). These findings underlined the very clear importance that public health messages should target people who are asymptomatic and those who notice a change in vision or are at higher risk of family history eye disease.

The prevalence of patients visiting optometric practice for routine eye examination in Australia was higher with increasing age. Although the prevalence of leading causes of some eye diseases and vision loss significantly increased among the elderly (Perruccio et al., 2007), the public awareness of the importance of routine eye examination should build among people in all age groups to encourage protecting their eyes. In contrast to the Australian finding, no significant association was found between the age and visiting the optometric practice for annual routine eye examination in the UK. It seems possible that this result may be due to increased awareness of eye care, recommendations given by optometrists and the good delivery of the eye care services in the UK, including reminding patients regularly by phone, text message or email when the next eye test is due.

Prior study has shown that lack of government-funded annual eye examination was correlated to 50% higher vision problems in low-income population compared to high-income population (Jin et al., 2013). Many residents in Australia and the UK are eligible for free eye examination (provided they meet certain criteria). Tylor (2004) suggested that there were many issues in the Australian community about delivery of eye care services. One of these issues was looking at how to increase public awareness of eye care and the importance of routine eye examination to reduce the impact of eye diseases.

Despite people being entitled for free regular eye examination, it has been found that some people are worried about having a regular eye test because they think they will be pressured into buying high cost spectacles (Shickle and Griffin, 2014). Therefore, it is important to change the public understanding of routine eye examinations in terms of detecting the early causes of preventable vision loss rather than only tests for refractive errors and sales of spectacles. Cost-saving is led by earlier treatment and detection of eye diseases rather than presenting with symptoms (Shickle et al., 2015).

The result showed that the second most common complaint after reading difficulty in Australia and the UK was change in spectacles prescription, which means patients returning to the optometrist with complaints about their glasses (Australia 24%, UK 19%). There are several possible causes for spectacle non-tolerance, including reasons related to spectacle dispensing, spectacle prescription, data entry error, pathology and binocular vision anomalies (Freeman and Evans, 2010). Age appears to be a factor in spectacles non-tolerance as the regression analysis revealed that the frequency of change in spectacles prescription complaint increased in the older age group. These results are in agreement with the findings Atchison et al. (2001) and Freeman and Evans (2010) that presbyopic people return to the optometric practice for non-tolerance complaint more than younger people. This result may be explained by the fact that people above 40 years old experience a gradual loss of accommodation over a period of approximately twenty years (Bennett and Rabbetts, 1998). Therefore, understanding the reasons of spectacles non-tolerance can help to improve both the practice and patient satisfactorily.

Regarding the interval of annual routine eye examination, the results found that most patients visited the optometrist every two years (30% in Australia, 27% UK). As mentioned earlier the frequency of routine eye examination depends on patient age and condition. However, one study found that having a routine eye examination every year was associated with better vision, especially for patients aged between 40 to 65 years (Li et al., 2013).

2.5 Conclusion

To answer the question of ‘why do we need to address the most common reasons for visiting optometric practices?’ This is because eye health information and data gathering about the most common vision problems are two of the most important aspects to provide efficient eye healthcare and identify eye health priorities. Reading difficulty was the most commonly reported complaint in Australia and the UK. In view of this finding it would be interesting to investigate the methods used in routine optometry practice for assessing near vision.

Study 2: Types of near vision charts used in optometric practice

The previous study findings confirmed that reading difficulty is one of the most common reasons for patients presenting for eye examination to an optometrist, and these results raise the importance of evaluating and standardising the near vision charts used in optometric practices. A reliable reading test provides results that help evaluate the treatment outcome of eye diseases (Stifter et al., 2005c). This study investigates which near vision charts are commonly used in optometry practice.

There are several types of reading charts used in research for scientific purposes, or in clinical practice, such as unrelated words reading test, continuous text (meaningful) reading tests, mixed contrast sensitivity tests, and computer-based reading tests. In addition, a variety of other different charts is currently used in daily clinical practice for reading performance measurements. The purpose of this survey was to establish what type of reading charts the optometrists routinely use in their clinics, and the outcome of this survey will be the first step towards preparing an inventory of all reading charts available in routine clinical practice and to compare their validity and repeatability, in order to obtain standardised and effective reading test measurements in daily clinical practice.

2.6 Methods

The data was collected through the distribution of an electronic (online) link to the optometrists. The link to a questionnaire was distributed using social media (Twitter, LinkedIn and Facebook) see Appendix 1. Ethical approval was obtained from Aston University Ethical Committee and the study adhered to the Declaration of Helsinki. Since the participants of optometrists for electronic questionnaire were recruited via social media it was difficult to know the response rate. Response rate cannot be determined because the number of participants who received and view the information is not clear (Schleyer and Forrest, 2000).

The practitioners were requested to indicate the type of practice they worked in and years of experience. The optometrists were asked to indicate the percentage of the patients usually seen in their clinic according to the patient age groups, which were categorised as pre-school children, school age children (5-16 years), pre-presbyopic age (16-44 years), presbyopic age (45-60 years) and over 60 years old.

In order to allow a comparison to the results obtained from the previous study the practitioners were asked to give their opinion on the most common chief complaints of patients seeking their clinic. For the main purpose of the questionnaire, the optometrists were asked to indicate the type of reading test charts and distance test chart they routinely used or that were available in their clinics. To analyse the data, descriptive statistics used (average, mean and percentage).

2.7 Results

Over a five-month period, 102 optometrists returned the questionnaires. The majority of the respondents worked in hospitals (47%), single independent clinics (23%) and (13%) large multiple chain stores (more than 30 sores). Less than 10% of the respondents were working in laser clinics, chain stores, or franchises. The practitioners' years of experience was (mean, 10 ± 9) years.

The respondents answered that the most common reasons for patients visiting their clinic were routine eye examination (65%) followed by reading difficulty complaint (51%). Figure 2.4 shows the percentages of the patients visiting optometrists for other reasons. From the chart in Figure 2.5, it can be seen that by far the greatest percentage of people visiting the optometric practices were pre-presbyopic (31%) and presbyopic (30%), followed by people aged over 60 years old (24%).

The results of this survey showed that the majority of the optometrists (65%) measure the distance visual acuity using Snellen chart (printed, backlit, projector or on a screen) while only 19% of them were using the logMAR chart.

The main question asked of the respondents was about the types of reading test charts they routinely used or were available in their clinics. The answers showed a wide range of charts being used; 65% of these charts were Times New Roman charts but of many different styles and often supplied by ophthalmic companies or by sponsors of eye conferences. The responses to this question reported by most of the optometrists stated answers such as "one chart provided by Spec Savers", "one chart provided by Nidek", "one chart provided by Precision Vision" and so on. Twenty-four per cent of reading charts reported were near Snellen reading chart while only 7% mentioned Jaeger reading chart. Of the 102 optometrists who completed the questionnaire, just two optometrists used Radner reading chart and one optometrist used the unrelated word Baily-Lovie reading chart.

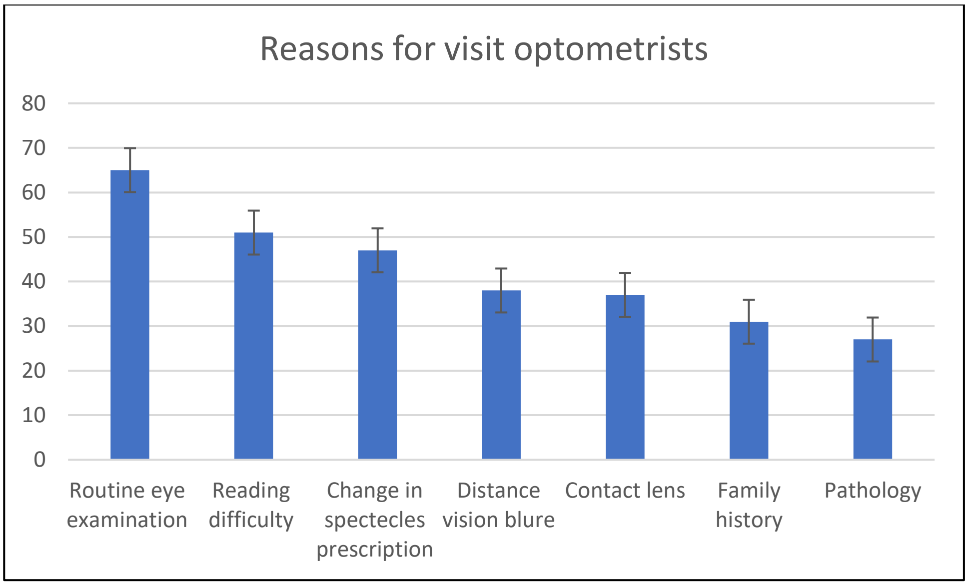


Figure 2.4: the optometrists' answers about the most common reasons for patients visiting their clinic

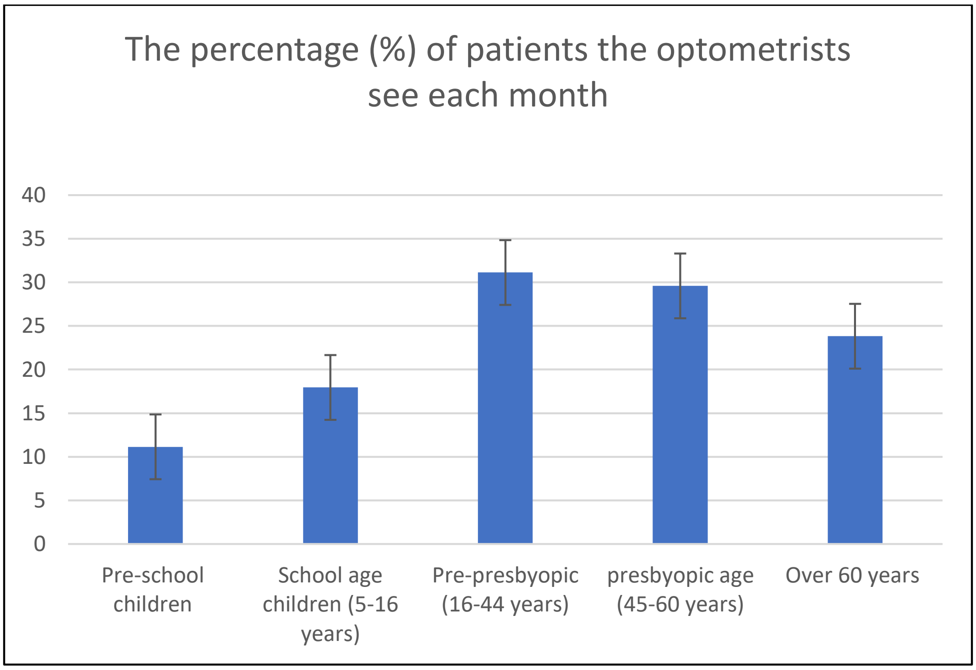


Figure 2.5: The number of patients seen by the optometrists each month according to their age group

2.8 Discussion

From the point of view of optometrists, the results of this survey confirmed that one of the most common reasons for visiting an eye clinic is a reading difficulty. This finding is in agreement with the previous study conducted in Australia and the UK. In spite of this finding, the use of the non-standardised reading charts is very common in clinical practice. A high percentage of the optometrists use reading charts that are created and printed by the eye care companies for advertising purposes.

The question about the type of distance test chart the optometrists used was included to establish and compare the extent of using the distance and near vision charts in terms of agreements between the optometrists. The results showed that the commonly used distance test chart was Snellen chart (65%). Although criticisms of the Snellen chart are widespread, it still seems to be the most favoured and agreed among optometrists. Limitations for using Snellen chart to measure distance visual acuity include the crowding effect, the effect of contour interaction, the number of the letters varies from line to line and the ratio of the letter size from one line to another is not constant (Jackson and Bailey, 2004). In addition, 24% of optometrists evaluated the reading performance by using Snellen reading chart and this type of chart also has not been developed in accordance to the recommendation of a standardised reading test chart.

The survey results reflected the real situation among the optometrists regarding the variety of reading test charts used in clinical practices today for measurement of reading performance. In contrast to distance visual acuity, there is no consensus regarding the charts or tests that should be used for measuring reading performance. In addition, the use of non-standardised reading charts (Times New Roman) is very common in clinical practice. These charts have not been developed according to recommendations for the standardised reading acuity charts (Colenbrander and Committee, 1988) and their repeatability and reliability are unknown. It is somewhat surprising that only three optometrists were using a reading chart developed according to reading chart recommendations, namely, the Radner and Bailey-Lovie reading chart. Standardised reading chart recommendation includes the use of logarithmic progression from one sentence to the next (Bailey and Lovie, 1976), same number of words in each sentence and the spaces between the letters and lines are proportional to the letter size. These results show the importance for the need of standardised reading tests.

Another important finding, as shown in Figure 2.5, was that a relatively high percentage of presbyopic people (30%) and people over 60 years old (24%) usually presented to the optometry clinic. As mentioned in the earlier, the ageing and growth of the world’s population is leading to increase the number of people with presbyopia (Bourne et al., 2017). Therefore, appropriate eye care and relevant eye examinations should be delivered to the people with presbyopia and this includes the use of reliable and repeatable reading performance test charts.

2.9 Conclusion

The consensus that reading difficulty is the most common complaint gives the standardised reading tests for reading performance measurements a significant value in clinical practice or research. The variety of reading test charts used today in clinical practices is strong evidence that there is no agreement regarding which reading chart should be used. Therefore, the next step of this project is to evaluate the available reading charts with regard to reliable and repeatable measurements in clinical practice.

Chapter 3

Reliability and repeatability of the reading performance metrics in different reading charts with young healthy subjects

3 Introduction

Reading is a highly complex task involving the integration of accurate eye movements, visual sensory input and higher cognitive aspects of comprehension (Latham and Whitaker, 1996). Reading ability has the largest influence on a person's daily life and functioning (Fujita et al., 2008). It is therefore becoming increasingly important to use standard clinical test to assess the impact of the vision on patient's quality of life. Reading performance measurement including an evaluation of the reading acuity and reading speed is a useful tool because it is simple to perform and provides a large amount of information.

The previous chapter findings concluded, there is no agreement regarding the charts or tests that should be used for measuring near reading acuity and function, resulting in unstructured development of reading charts. Many types of texts, charts or even newspapers are currently used in daily clinical practice for measuring reading performance and near visual acuity. However, to obtain a comparable outcome in daily clinical practice and research, evaluation and standardization should be undertaken on every clinical test that assesses reading performance.

The use of non-standardized reading charts is very common in clinical practice; these are created and printed by practitioners or distributed by eye care companies for the purpose of advertising their services. These charts have not been developed according to recommendations for the standardized reading acuity charts and their repeatability and reliability have also not been evaluated (Colenbrander and Committee, 1988).

The standardized reading tests used in clinical practice or in research studies can increase the validity of prediction of the real-world performance. There are several types of reading tests used in research, and in clinical practice. However, the majority of previous studies have described and evaluated only three charts of continuous text reading tests: IReST in one print size, the Radner, and the MNread. No studies have made a direct comparison between these continuous text reading charts (Brussee et al., 2014). Number of studies available regarding the validity of the existing reading tests, however, the methodology, the content of the study design and the language used varied. On the other hand, some of the design aspects of some studies are missing or are not clear, which makes the comparison between the outcomes difficult.

The purpose of this study is to statistically analyze and compare the inter-chart reliability and test-retest repeatability of the reading charts (MNread, Radner, Colenbrander and IReST) for measuring reading performance metrics in normally sighted subjects. In addition, to compare the available continuous text reading tests properties with unrelated words on the Bailey-Lovie reading chart. Times New Roman (TNR) chart which is distributed to clinical practices by Ciba Vision (Alcon, Fort Worth, Texas, USA) was included in the analysis due to the previous findings of frequent use of this type of chart in the UK.

In the evaluation of the reading performance, the testing parameters of various reading tests should provide minimal variability and differences in the outcomes to indicate equality for clinical examination. Therefore, one aim of this study is to investigate whether the clinical results of the reading performance metrics with different reading test charts are affected by the choice of the charts.

3.1 Materials and methods

In the present study at two testing sessions (two to three weeks apart), 29 normally sighted participants were recruited. Exclusion criteria include subjects for whom English was not their first language, those with ocular diseases, use of medication that may influence cognitive ability, presence of binocular vision anomalies and reading disorders. Subjects with best corrected visual acuity at distance and near worse than 0.1 logMAR were excluded. Older patients who already had presbyopia or had signs of early presbyopia were excluded. Ethical approval was obtained from Aston University Ethical Committee (see Appendix 3) in accordance with Helsinki Declaration and all the subjects gave their consent to participate in the study. All tests were performed monocularly with the dominant eye wearing the optimal refractive correction.

Corrected distance visual acuity was measured using the Early Treatment Diabetic Retinopathy Study (ETDRS) chart displayed by software CSO vision chart v2.6.0 on an LCD screen at distance of 4m. Near visual acuity was measured using ETDRS near vision chart (Precision Vision) at 40cm. Refraction was performed for all participants. To ensure the subjects had no accommodation anomalies, amplitude of accommodation was measured using the Royal Air Force (RAF) rule.

3.1.1 Materials

Reading performance test charts available in the English language and used in this study were: MNread, Radner, Colenbrander, IReST, Times New Roman and the Bailey-Lovie word reading chart.

The Times New Roman chart type included in this analysis was distributed to clinical practices by Ciba Vision. The Ciba Vision reading chart contains seven paragraphs with the number of words ranging from 5 to 80. The smallest print size in the Ciba Vision reading chart is N6 (Fig 3.1).

Aston University

Illustration removed for copyright restrictions

Figure 3.1: Ciba Vision Times New Roman reading chart

For all the reading charts, except for the Times New Roman and IReST, print sizes adopt a logarithmic progression for simultaneous determination of the reading speed as well as reading acuity and critical print size.

IReST
International Reading Speed Texts

In a small town a greengrocer had opened a shop that was located above a deep cellar. Every night, mice came in droves out of this cellar into the shop. They ate apples and pears, grapes and nuts and did not spare the vegetables and potatoes either. No goods that were in the shop were safe from the small intrusive rodents between midnight and sunrise. As long as there was noise in the streets at night and cars were driving by, the mice still stayed quietly in the cellar. But as soon as the old clock on the town hall had struck midnight and it became quiet in the street, they came out in droves, enjoyed the sweet fruits and celebrated real feasts, whose remains filled the owner with despair every morning when he entered the shop. So he tried to protect himself against the mice. At first he set up traps all over the shop.

number of text: 1
name of text: Mice
performance category: AB
number of words: 156
number of syllables: 205
number of characters: 662
reading time in seconds (mean ± SD): 40.4 ± 6.2
reading speed (mean ± SD) words/minute: 236 ± 29

Figure 3.2: IReST reading test (text number 1)

The IReST reading test measures reading performance under natural conditions as the texts are provided in a size of 1.7mm (newspaper print size) and use one print size (Fig 3.2). IReST contains 10 long paragraphs of text that are equal in difficulty, length and linguistic complexity. The number of words in each paragraph ranges from 136 to 165 (in the English version), see section 1.5.5.

The MNread acuity card consists of 19 sentences with 10 to 14 words per sentence displayed in three rows with controlled spatial layout and word frequency. For measuring the reading speed, each sentence should consist of 10 words assuming an average English word length of six characters (Legge, 2006). The MNread chart print sizes range from +1.3 to -0.5 logMAR (Fig 3.3), see section 1.5.2.

The Radner reading chart consists of 14 sentences with 14 words per sentence presented in three lines with controlled position of the words and equal number of characters (Fig 3.4). The print sizes in the Radner chart range from +1.1 to -0.2 logMAR, see section 1.5.3.

The Colenbrander reading chart is a type of continuous text reading test with high contrast continuous text print sizes (Fig 3.5). The Colenbrander chart uses 14 pairs of sentences with equal length and spacing and with print sizes, see section 1.5.4.

The Bailey-Lovie unrelated words reading chart evaluates the reading performance with unrelated words rather than meaningful sentences. It is viewed at 25cm with print size ranging from 0.0 to 1.6 logMAR in 0.1 log unit steps with logarithmic progression (Fig 3.6). However, in this study, the Bailey-Lovie reading chart was presented on a reading stand with a reading distance of 40cm and the correction distance factor was used. In the Bailey-Lovie reading chart, each step represents 0.1 log units (Bailey and Lovie, 1980). The correction factor was determined by the number of steps between 40cm and 25cm; reading distance 40cm is two steps further than 25cm. In the current study, participants read the logMAR row at 40cm and then a correction factor of 0.2 log units was subtracted to achieve the reading acuity score at 40cm, see section 1.5.1.

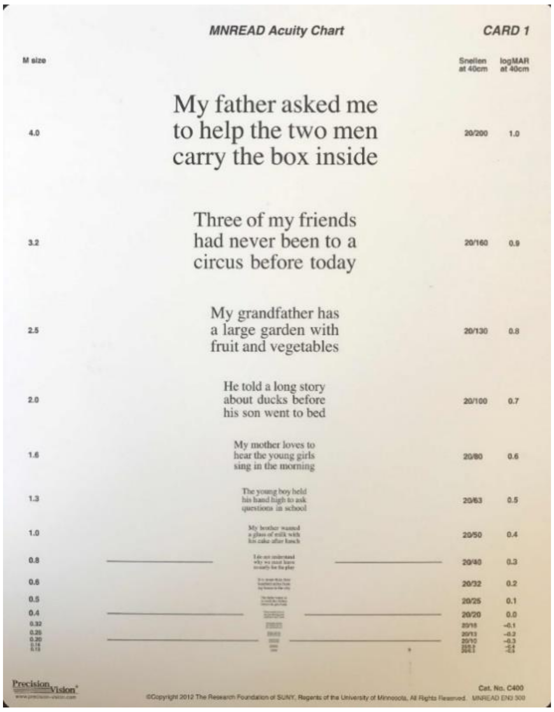


Figure 3.3: MNread acuity chart.

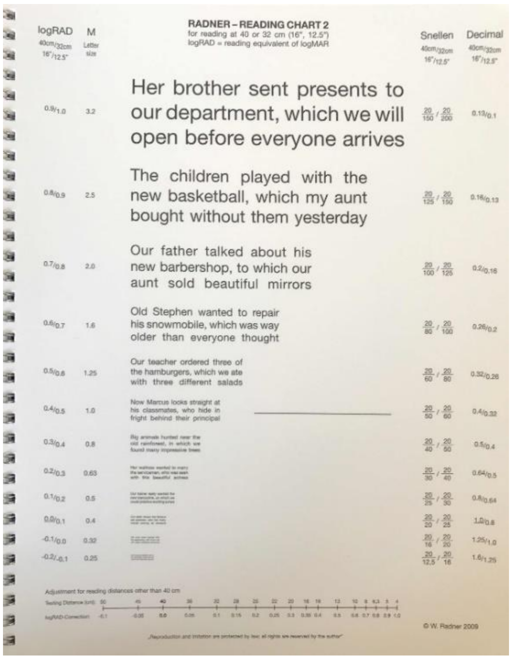


Figure 3.4: Radner reading chart

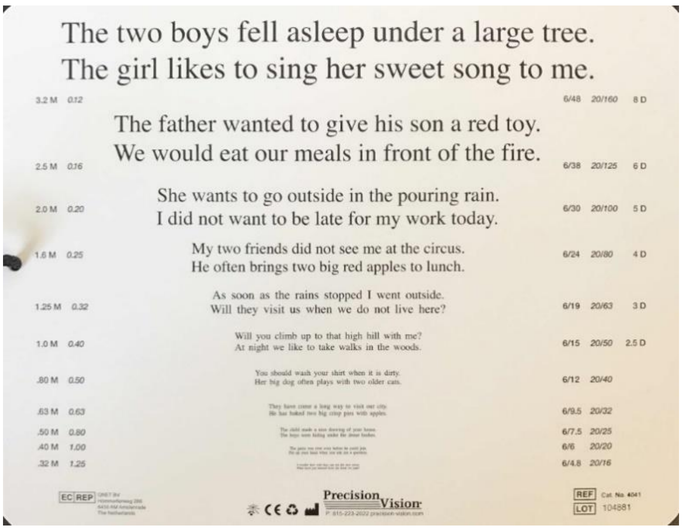


Figure 3.5: Colenbrander reading chart

Figure 3.6: Bailey-Lovie unrelated words reading chart. With permission of https://www.researchgate.net

3.1.2 Measurements procedure

All the reading tests use Times New Roman font, which is a serif font, except the Radner, which uses the sans serif Arial font. All reading tests were presented on a reading stand at a distance of 40cm determined by using a centimeter ruler. To ensure constant viewing distance, a headrest for the forehead was used. The luminance from all the reading test charts was between 120- 135 cd/m². The luminance was measured by photometer (AP-881D Digital LUX Meter).

To enhance and achieve comfortable reading performance, a daylight lamp was used during the test sessions. The lamp was positioned in front of the reading test chart instead of placing it over the participant's shoulder to avoid the possibility of body shadow that could affect the chart illumination (Fig 3.7).

The number of the paragraphs or sentences per reading chart varies: IReST (10), Ciba Vision Times New Roman (7), Colenbrander (14) sentences, Radner (14), MNread (19) and Bailey-Lovie (17). All the reading tests were performed monocularly with the dominant eye. The dominant eye was identified as the eye used for sighting when the participant looked at a distant object through a ring while both eyes were open (Shneor and Hochstein, 2008).

Each reading chart was covered by a piece of paper and the participants were asked to read the sentences as soon as the examiner removed the cover. The participants were instructed to read the sentences one after the other, loudly and accurately, as quickly as possible without correcting themselves if they made errors and without repeating the words or stopping in between. Reading errors were counted using the following criteria: 1- the word read incorrectly, 2- an added word, 3- an omitted word, 4- a repeated word.

During the session, a short break of around three minutes was taken before starting the next reading test to eliminate any possible fatigue effect. The presentation sequence of the six reading tests was randomized to avoid learning effects. For the IReST reading tests, the participants were asked to read all 10 paragraphs. Testing procedures were monitored with audio recording to measure the reading time for each sentence and reading errors. Reading performance metrics measurements were carried out after the testing session, from the audio recordings. GoldWave Inc (St. John’s, Newfoundland, Canada) software was used for playing the audio recordings. All the test procedures and measurements were performed by one examiner. The second examinations were performed two to three weeks later. For subjects wearing spectacles, the same spectacles were used in both testing sessions.

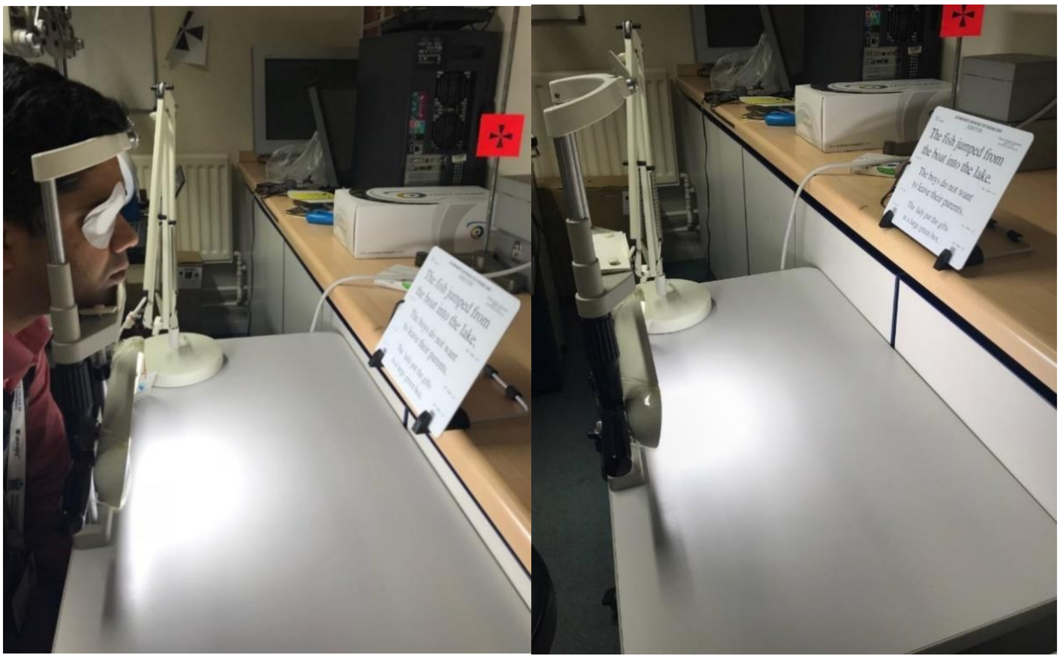


Figure 3.7: Experiment’s setting procedure. Reading test charts were placed on holder parallel to the participant’s face. The non-dominant eye was occluded (Baashen, 2020).

3.1.3 Calculation of reading performance metrics

Following two testing sessions, the reading performance metrics outcomes were described as follows: reading acuity (RA), maximum reading speed (MRS), critical print size (CPS) and logRAD/logMAR ratio.

For the IReST reading test, just one metric was measured, which was the average reading speed, because the IReST is a one print size chart.

Reading acuity was determined as the smallest print size that could be read, taking the reading errors into account. The test end point for reading acuity was defined as reading more than 80% of the words in the sentences accurately (Stifter et al., 2005). Reading acuity is expressed in terms of logRAD, which is the reading equivalent of the distance logMAR. For the five reading tests (IReST, Times New Roman, Colenbrander, MNread and Bailey- Lovie chart), the number of errors for each sentence was counted in number of words read incorrectly. For the Radner test, to calculate reading acuity, the number of errors was counted in number of syllables of incorrectly read words, as recommended by the company. Each syllable in a Radner chart has a value of 0.005, while for calculating MNread test reading acuity, each word in the chart has a value of 0.01 logMAR. Unfortunately, in Colenbrander and Bailey-Lovie reading charts, the value of each word was not clearly mentioned by the developers. Therefore, the reading acuity was measured as the smallest print size attempted without taking reading errors into account.

Radner reading acuity (logRAD) = smallest print size attempted + (total number of syllables of incorrectly read words X 0.005)

MNread reading acuity (logRAD) = smallest print size attempted + (total number of incorrectly read words X 0.01)

Reading speed (RS) is measured in words per minutes (wpm) and calculated, for each sentence, as the number of words read correctly divided by the exposure time taken to read that sentence.

$$\text{RS (wpm)} = \frac{\text{Number of the words read correctly in a sentence}}{\text{time in second need to read the sentence}} \times 60$$

Critical print size was determined as the steepest decline slope on the graph of plotted reading time in wpm per sentence for each chart and each subject was used (Burggraaff et al., 2010b). Maximum reading speed was the average reading speed of the sentences corresponding to and larger than CPS. To compare the near and the distance visual acuities, logRAD/logMAR ratio was calculated in a percentage as follows:

$$\text{logRAD/logMAR ratio (\%)} = \frac{1-\text{logRAD}}{1-\text{logMAR}} \text{ X } 100$$

3.1.4 Statistical analysis

The required sample size for this study was calculated with a statistical power analyzer (G*power 3.0.10) and the following parameters used: $1\text{-}\beta = 0.80$, $\alpha = 0.05$, paired t-test. The calculation resulted in a requirement of sample size of at least $n = 27$ subjects. Thus, this study required 30 subjects (allow 10% for drop out for repeat testing)

The statistical analysis was performed using Statistical Package for Social Sciences (SPSS) version 23 and Microsoft Excel 2013 for Windows. The data was normally distributed (Shapiro-Wilks test $p > 0.05$). The differences of the reading performance metrics measurements between the two testing sessions, as well as the number of reading errors, were analyzed for significance by means of t-test. The cut-off level for statistical significance was $p<0.05$, two-tailed t test.

To measure the agreements of the reading performance metrics between the two testing sessions, Bland-Altman analysis was used (Bland and Altman, 1986). In addition to Bland-Altman analysis, Pearson correlation coefficient was calculated to assess the test-retest reliability between the first and second visits. Reliability of the reading charts was determined by calculating the Cronbach's alpha coefficient.

3.2 Results

Reading performance data of 29 participants were included in the analysis, after excluding three subjects, one subject with reading problem (dyslexia), the second subject having a binocular vision problem and the third subject having some difficulty in pronunciation of the English words. The participants in this study were recruited from among Aston University students. Subjects' age ranges from 18 to 31 years (mean, 20.4 ± 3.7). There were 20 with right eye dominance and nine with left eye dominance. Distance visual acuities were (mean, -0.11± 0.08) logMAR and near visual acuities were (mean, -0.06 ± 0.08) logMAR. The mean value of the spherical equivalent of the refractive error was (-1.00 ± 1.50) D. The amplitude of accommodation for all the participants was within a normal range for their age (8.9D ± 2.4D).

3.2.1 Repeatability

3.2.1.1 Continuous-text reading tests (Radner, MNread and Colenbrander)

In the current study, repeatability of a test refers to a test-retest reliability which measures the variability of the measurements obtained by one examiner on the same participants and same methods repeatedly with time interval two to three weeks under the same condition. Colenbrander, Radner and MNread are types of short sentences continuous text reading tests arranged in logarithmic progression.

Table 3.1: Test-retest reliability between the first and the second testing session. Pearson correlation coefficients n =29

Reading metrics	**Colenbrander** (***r*** **value,** ***p*** **value)**	**Radner** (***r*** **value,** ***p*** **value)**	**MNread** (***r*** **value,** ***p*** **value)**
Reading acuity (logRAD)	$r = 0.772$, $p < 0.0001$	$r = 0.630$, $p = 0.0002$	$r = 0.816$, $p < 0.0001$
Maximum reading speed (wpm)	$r = 0.550$, $p = 0.001$	$r = 0.895$, $p < 0.0001$	$r = 0.856$, $p < 0.0001$
Critical print size (logMAR)	$r = 0.297$, $p = 0.12$	$r = 0.424$, $p = 0.02$	$r = 0.509$, $p = 0.005$
LogRAD/LogMAR ratio (%)	$r = 0.755$, $p < 0.001$	$r = 0.582$, $P = 0.001$	$r = 0.701$, $p < 0.0001$

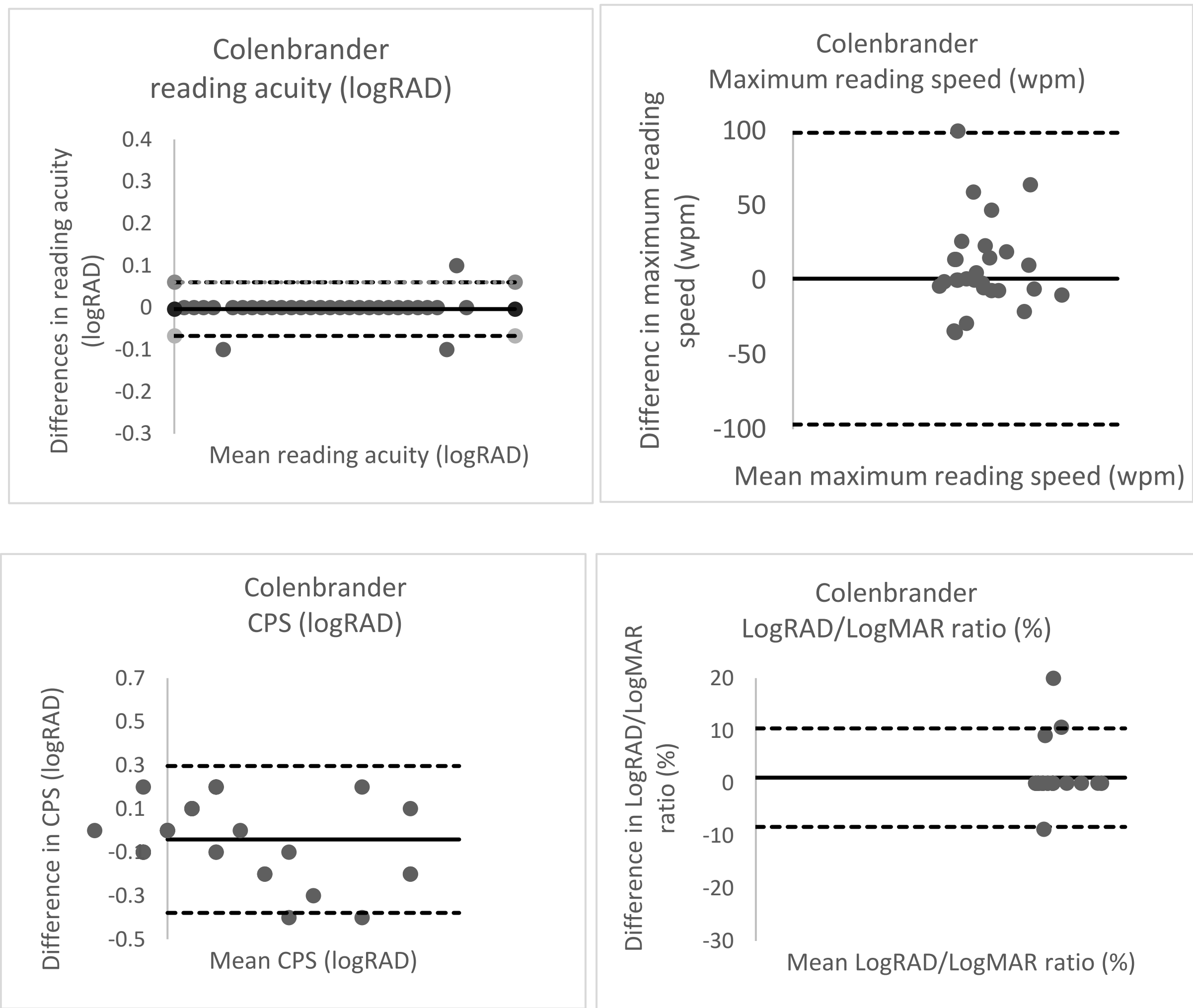
The test-retest correlation for the three reading tests was high for reading acuity. However, this correlation was higher with the MNread chart (r = 0.816) followed by Colenbrander (r =0.772) and then Radner (r = 0.630). For maximum reading speed, the correlation between the two testing sessions in the Colenbrander chart (r = 0.550) was lower than the MNread and Radner charts (r = 0.895, r = 0.856). For critical print size, no correlation was found in the Colenbrander chart while, for the Radner and MNread charts, moderate correlation was found (r = 0.509, r = 0.424). For LogRAD/LogMAR ratio, the correlations were higher for Colenbrander and MNread (r =0.755, r = 0.701) but low for the Radner chart (r = 0.582). As shown in Table 3.1, all the correlations were tested for statistical significant P value < 0.05. In overall, the MNread chart showed a good test-retest correlation for all reading performance metrics compared to Radner and Colenbrander.

In the Radner reading chart, statistically significant difference between the two testing sessions was found with respect to the reading acuity (p = 0.008) and maximum reading speed

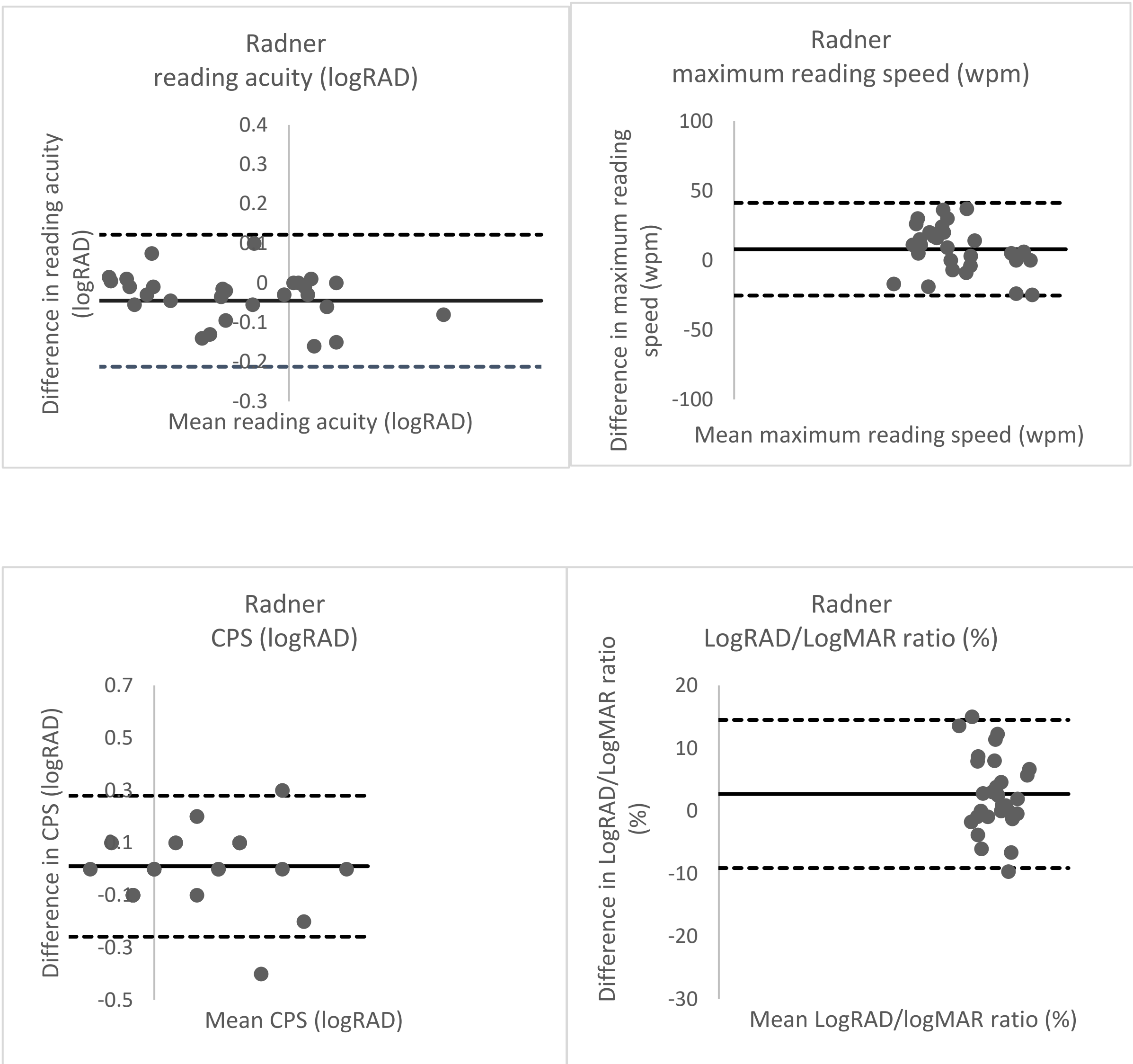
(p = 0.01). These differences were: LogRAD: -0.045 ± 0.085, 95% confidence interval CI: (-0.077; -0.013); MRS (wpm): 7.931± 16.981, 95% confidence interval CI: (1.472; 14.390). For the Colenbrander and MNread charts, no significant differences between the two testing sessions were found for all the reading performance metrics.

As the Pearson correlation measured only the strength of the relation between the testing sessions, Bland-Altman analysis (Bland and Altman, 1986) was also performed to evaluate the agreement between the testing sessions for the three reading charts with respect to all reading performance metrics. In the Colenbrander and Radner charts, Bland-Altman analyses evaluated a good repeatability between the two testing sessions for MRS, CPS and LogRAD/LogMAR ratio. However, low level of agreement was resulted for reading acuity. In MNread, Bland-Altman analysis showed a good agreement for all the reading performance metrics except for CPS where the level of agreement was low (Fig 3.8).

a)



b)


Radner
reading acuity (logRAD)
Difference in reading acuity (logRAD)
Mean reading acuity (logRAD)
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Radner
maximum reading speed (wpm)
Difference in maximum reading speed (wpm)
Mean maximum reading speed (wpm)
100
50
0
-50
-100
Radner
CPS (logRAD)
Difference in CPS (logRAD)
Mean CPS (logRAD)
0.7
0.5
0.3
0.1
-0.1
-0.3
-0.5
Radner
LogRAD/LogMAR ratio (%)
Difference in LogRAD/LogMAR ratio (%)
Mean LogRAD/logMAR ratio (%)
20
10
0
-10
-20
-30

C)

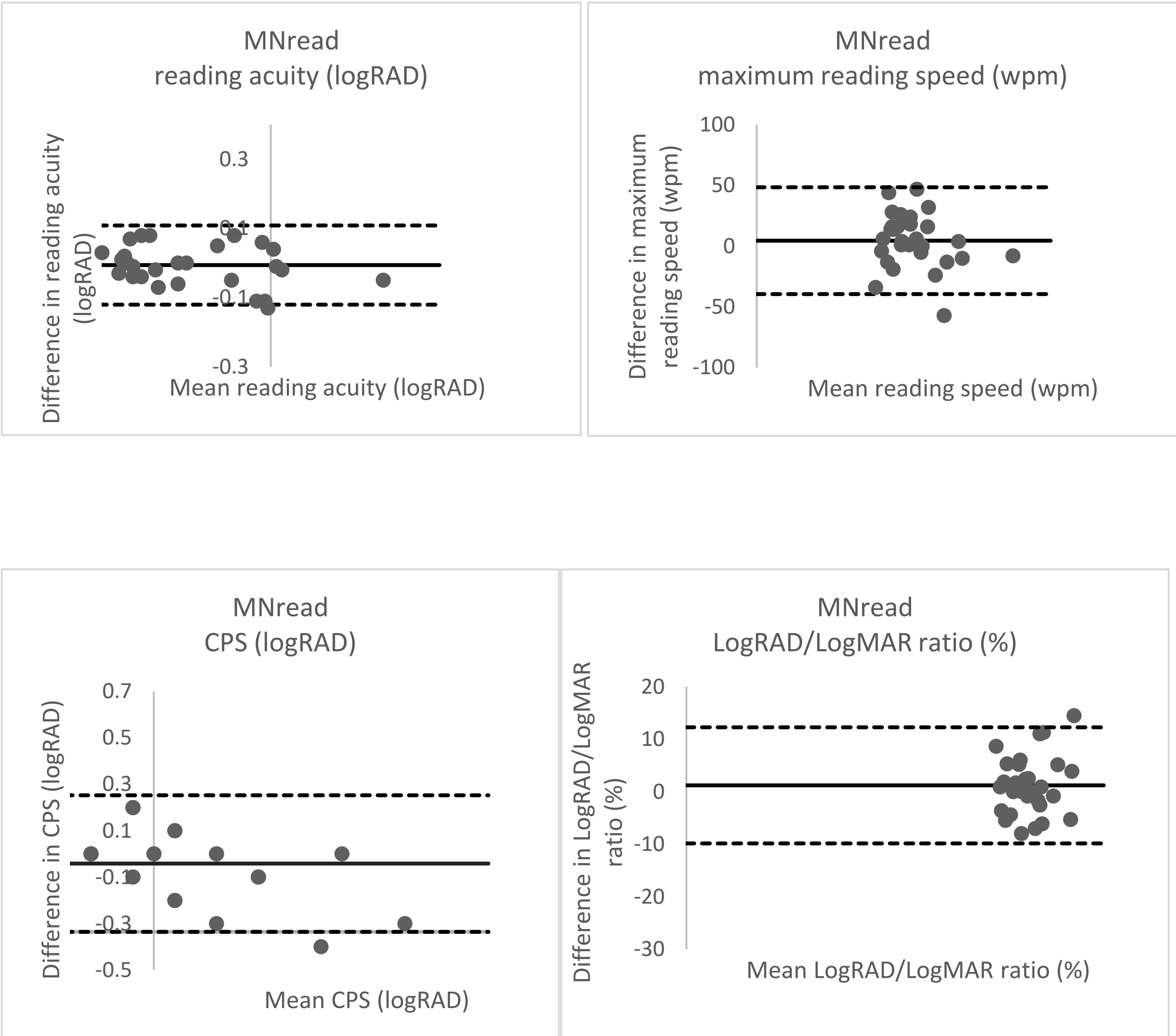


Figure 3.8: Test-retest reliability: Bland-Altman plot of the differences between the two testing sessions (second minus first). In **(a)** Bland-Altman plots for reading performance metrics in the Colenbrander Chart showed high agreement between the two testing sessions for MRS, CPS and LogRAD/LogMAR ratio, but not for RA; **(b)** Bland-Altman plots for reading performance metrics in the Radner chart showed a low agreement in reading acuity; **(c)** Bland-Altman plots for the MNread chart showed a low agreement in CPS, but high agreement for RA, MRS and logRAD/logMAR ratio (*n* = 29).

3.2.1.2 Repeatability of Bailey-Lovie unrelated words reading chart

In the Bailey-Lovie word chart, Pearson correlation coefficients showed that MRS (r = 0.572), CPS (r = 0.580) and LogRAD/LogMAR ratio (r = 0.608) were moderately correlated between the two testing sessions. However, the correlation was higher for reading acuity (r = 0.795). The differences between the two visits were statistically significant for reading acuity (p = 0.001) and LogRAD/LogMAR ratio (p = 0.003). These differences were: LogRAD: -0.045 ± 0.068, 95% confidence interval CI: (-0.071; -0.019); and LogRAD/LogMAR ratio (%): 7.931± 16.981, 95% confidence interval CI: (1.391; 6.261).

The Bland-Altman agreement in measuring critical print size was low, but higher agreements between the two sessions were found for RA, MRS and LogRAD/LogMAR ratio (Fig 3.9)

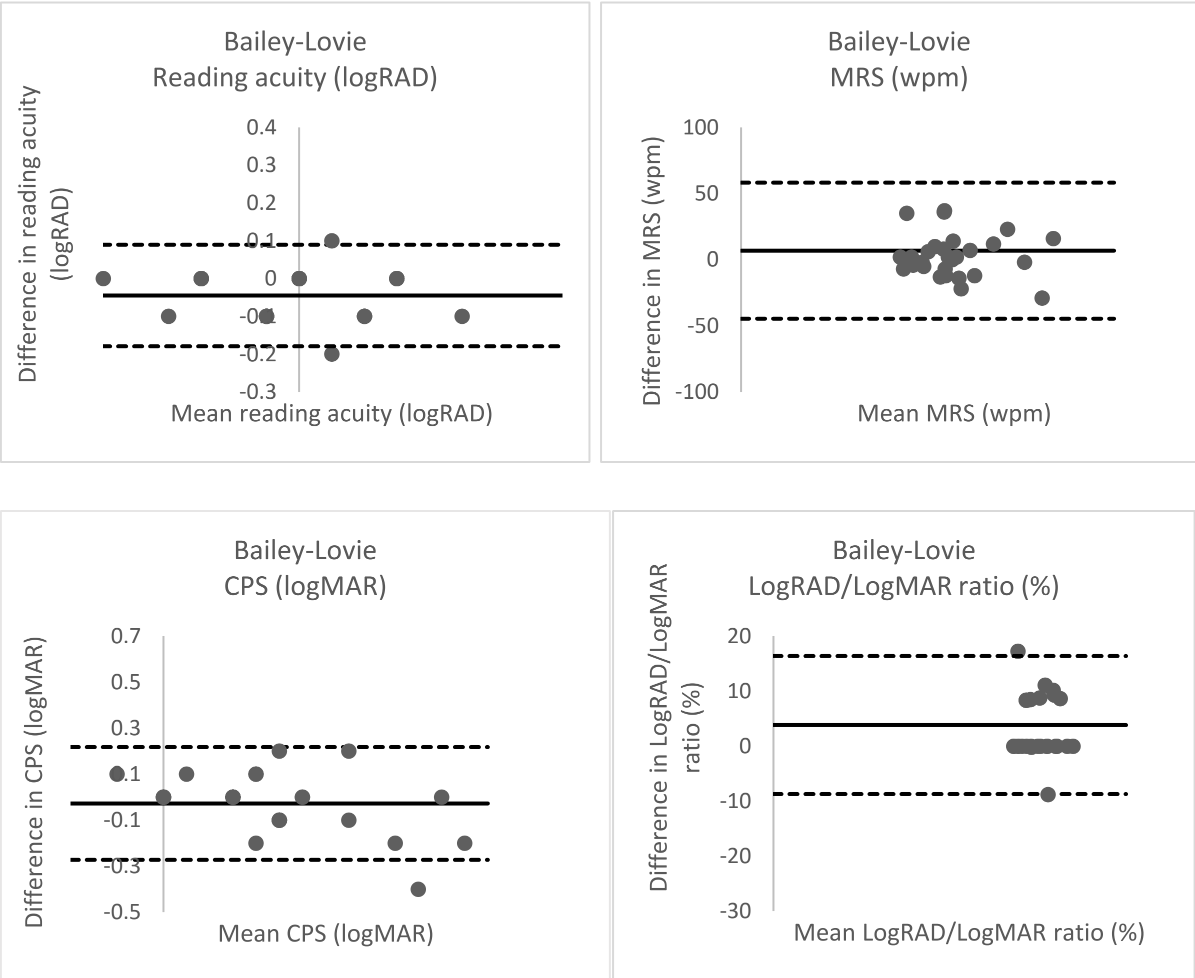


Figure 3.9: Test-retest reliability Bailey-Lovie words chart: Bland-Altman plots of the differences between the two testing sessions (second minus first) for RA, MRS, CPS and LogRAD/LogMAR ratio. Higher agreements were found for all reading performance metrics, except for CPS where the agreement was low

3.2.1.3 Repeatability of Times New Roman chart (Ciba Vision reading card)

In the Times New Roman chart, all the participants could easily read the smallest print size, N6, in the two testing sessions. Therefore, the reading acuity was exactly the same in the two visits (N6). The test-retest measurements of MRS were moderately correlated (r = 0.687) but the Bland-Altman agreements between them were low. Very weak correlation and agreement between the two testing sessions were found for measuring CPS (r =0.272).

3.2.1.4 Repeatability of International Reading Speed test (IReST)

The only reading performance parameter tested for the IReST chart was the average reading speed, as the chart has only one print size. The measurements of the average reading speed of the 10 long IReST paragraphs were strongly correlated between the two testing sessions (r = 0.918). However, the difference in average reading speed was statistically significant (p= 0.0002). This difference was wpm: 11 ± 15, 95% confidence interval CI: (6.0; 17.722). Bland-Altman evaluated a good agreement in the repeatability of average reading speed measurements. Table (3.2) summarizes the repeatability results of the reading performance metrics for the six reading charts.

Table 3.2: Summary of the repeatability results of the reading performance metrics for the six reading charts.

Reading chart	Reading performance metrics	Correlation coefficient (r value)	Differences in measurements (t test) *p* value	Bland-Altman agreement
Colenbrander	RA (logRAD)	0.772	0.572	[0.06; -0.07]
	MRS (wpm)	0.550	0.920	[98; -96]
	CPS (logMAR)	0.297	0.206	[0.29; -0.37]
	LogRAD/LogMAR ratio (%)	0.755	0.239	[10.4; -8.3]
Radner	RA (logRAD)	0.630	0.008	[-0.01; -0.07]
	MRS (wpm)	0.895	0.01	[41; -25]
	CPS (logMAR)	0.424	0.687	[0.25; -0.27]
	LogRAD/LogMAR ratio (%)	0.582	0.022	[14.5; -9.1]
MNread	RA (logRAD)	0.816	0.593	[0.10; -0.12]
	MRS (wpm)	0.856	0.291	[48; -39]
	CPS (logMAR)	0.509	0.148	[0.25; -0.33]
	LogRAD/LogMAR ratio (%)	0.701	0.257	[12.2; -9.8]
Bailey-Lovie	RA (logRAD)	0.795	0.001	[0.08; -0.17]
	MRS (wpm)	0.572	0.172	[58; -44]
	CPS (logMAR)	0.580	0.244	[0.21; -0.27]
	LogRAD/LogMAR ratio (%)	0.608	0.003	[6.2; 1.4]

Times New Roman	MRS (wpm)	0.687	0.094	[80; -0.58]
	CPS (N notation)	0.272	0.169	[5.09; -5.23]
IReST	Average reading speed (wpm)	0.918	0.0002	[42; -18]

3.2.2 Reliability

In the current study, the terms reliability or inter-chart reliability refer to the internal consistency of the reading performance metrics measurements obtained from each reading chart at the two testing sessions. In the Colenbrander reading chart, the reliability analysis for RA, MRS and LogRAD/LogMAR ratio resulted in good Cronbach's alpha coefficients, while in measuring for the CPS the reliability was very poor (α = 0.456). In the Radner reading chart, higher reliability results were found for measuring MRS (α = 0.939) followed by RA (α = 0.764) and LogRAD/LogMAR ratio (α = 0.735). Poor reliability result was found for measuring CPS in the Radner reading chart (α = 0.581). In the MNread reading chart, the reliability for CPS measurements was lower than those for RA, MRS and LogRAD/LogMAR ratio. Interestingly, the Bailey-Lovie word reading chart show good reliability results for all reading performance metrics measurements. The reliability in reading speed measurements by IReST was very high (α = 0.951). Measuring the reliability analysis for the Times New Roman chart with respect to reading acuity measurements was difficult, because all the participants could read the smallest print size, N6. Therefore, the variables had zero variance. However, for CPS measurements, Times New Roman revealed a very poor reliability (α = 0.429). Table (3.3) summarizes the reliability results of the reading charts.

Table 3.3: The Cronbach's alpha coefficient results for evaluating the reliability of the reading chart

Reading chart	Reading performance metrics	Cronbach's Alpha	Reading chart	Reading performance metrics	Cronbach's Alpha
Colenbrander	RA (logRAD)	0.860	**Bailey-Lovie**	RA (logRAD)	0.881
	MRS (wpm)	0.697		MRS (wpm)	0.722
	CPS (logMAR)	0.456		CPS (logMAR)	0.717
	LogRAD/LogMAR ratio (%)	0.857		LogRAD/LogMAR ratio (%)	0.750
Radner	RA (logRAD)	0.764	**TNR**	MRS (wpm)	0.764
	MRS (wpm)	0.939		CPS (N notation)	0.429
	CPS (logMAR)	0.581			
	LogRAD/LogMAR ratio (%)	0.735	**IReST**	Average reading speed (wpm)	0.951
MNread	RA (logRAD)	0.894			
	MRS (wpm)	0.921			
	CPS (logMAR)	0.646			
	LogRAD/LogMAR ratio (%)	0.811			

3.2.3 Comparison of the reading errors between the charts

The average of the total reading errors between the different reading charts in the two testing sessions is given in Table (3.4). For the IReST, the total number of errors made in all ten long paragraphs was counted. That resulted in a greater number of errors in the IReST chart compared to the other short sentences reading charts. The participants made fewer reading errors in the second testing session compared to the first testing session, because of the much greater reading practice effect. In the first testing session, the participants made significantly more reading errors in the Radner reading chart compared to Colenbrander (p = 0.02), Bailey-Lovie (p= 0.0001) and Times New Roman charts (p = 0.001). No significant differences in the number of reading errors were found between Radner and MNread. In addition, the number of errors in the MNread chart was significantly higher than the Bailey-Lovie (p < 0.05) and Times New Roman (p = 0.01) charts.

Table 3.4: Number of reading errors per reading chart

Test session	**Colenbrander**	**Radner**	**MNread**	**Bailey-Lovie**	**Times New Roman**	**IReST**
First (mean± SD)	3.62 ± 2.79	6.21± 4.58	4.68± 3.46	2.20± 2.23	2.48± 2.42	13.20± 7.97
Second (mean± SD)	4.07 ± 3.27	3.72± 2.87	3.44± 3.51	2.58± 2.24	3.21± 2.45	10.41± 6.85

3.2.4 Comparison of reading performance outcomes between reading test charts

To test whether the reading performance metrics were equal between the reading test charts with respect to the normally distributed data (Shapiro-Wilks test p >0.05), Pearson correlation coefficient and Bland-Altman analysis by determining 95% limit of agreement were performed to evaluate the agreement between the test chart outcomes in the first testing session. Also, t- test was used to compare the mean differences of the reading performance outcomes between the reading charts. P < 0.05 was considered statistically significant. Since the TNR chart is not arranged in a logarithmic progression, the maximum reading speed was the only parameter compared with the other reading test charts.

Table 3.5: The correlation and differences together with Bland-Altman limit of agreement between the reading charts within the same test session. The differences in the reading performance metrics between the charts showed a variation of the results.

Reading charts	***Reading performance metrics***	***Correlation coefficient (r value)***	***Differences in measurements t- test p value (mean ± SD)***	***Bland-Altman limits of agreement***
Colenbrander + Radner	RA (logRAD)	0.0509	*P*= 0.028 (-0.04± 0.01)	[0.14; -0.2]
	MRS (wpm)	0.818	*P*= 0.000 (73± 26)	[121; 23]
	CPS (logMAR)	0.290	*P*= 0.141 (0.045± 0.159)	[0.69; 0.106]
	LogRAD/LogMAR ratio (%)	0.467	*P*= 0.112 (2.10± 6.87)	[15.3; -11.2]
Colenbrander+ MNread	RA (logRAD)	0.292	*P*= 0.791 (-0.0027± 0.055)	[0.2; -0.2]
	MRS (wpm)	0.438	*P*= 0.000 (110± 40)	[188; 32]
	CPS (logMAR)	0.444	*P*= 0.550 (0.02±0.18)	[0.37; -0.3]
	LogRAD/LogMAR ratio (%)	0.628	*P*= 0.609 (-0.53± 5.54)	[10.3; -11.3]

Table 3.5: (continued)

Reading charts	***Reading performance metrics***	***Correlation coefficient (r value)***	***Differences in measurements t- test p value (mean ± SD)***	***Bland-Altman limits of agreement***
Colenbrander + Bailey-Lovie	RA (logRAD)	0.704	*P*= 0.000 (0.1± 0.84)	[0.06; -0.26]
	MRS (wpm)	0.614	*P*= 0.000 (140± 36)	[210; 96]
	CPS (logMAR)	0.272	*P*= 0.115 (-0.05± 0.24)	[0.30; -0.41]
	LogRAD/LogMAR ratio (%)	0.364	*P*= 0.000 (7.70± 8.03)	[23.4; -8.2]
Radner+ MNread	RA (logRAD)	0.702	*P*= 0.00 (0.065± 0.08)	[0.4; -0.09]
	MRS (wpm)	0.891	*P*= 0.001 (-14± 19)	[23; -51]
	CPS (logMAR)	0.321	*P*= 0.914 (0.003±0.17)	[0.33; -0.33]
	LogRAD/LogMAR ratio (%)	0.658	*P*= 0.000 (4.25± 5.39)	[6.3; -14.8]

Table 3.5: (continued)

Reading charts	*Reading performance metrics*	*Correlation coefficient (r value)*	*Differences in measurements t- test p value (mean ± SD)*	*Bland-Altman limits of agreement*
Radner+ Bailey-Lovie	RA (logRAD)	0.647	*P*= 0.001 (-0.06± 0.92)	[0.12; -0.24]
	MRS (wpm)	0.767	*P*= 0.000 (67± 24)	[115; 19]
	CPS (logMAR)	0.373	*P*= 0.001 (-0.1± 0.15)	[0.19; -0.39]
	LogRAD/LogMAR ratio (%)	0.524	*P*= 0.000 (5.64± 7.11)	[19.5; -8.2]
Bailey-Lovie+ MNread	RA (logRAD)	0.420	*P*= 0.000 (-0.09± 0.10)	[0.09; -0.29]
	MRS (wpm)	0.632	*P*= 0.000 (29± 21)	[70; -12]
	CPS (logMAR)	0.510	*P*= 0.025 (-0.06±0.17)	[0.27; -0.39]
	LogRAD/LogMAR ratio (%)	0.324	*P*= 0.000 (8.24± 10.66)	[29.1; -10.0]

Table 3.5: (continued)

Reading charts	*Reading performance metrics*	*Correlation coefficient (r value)*	*Differences in measurements t- test p value (mean ± SD)*	*Bland-Altman limits of agreement*
Colenbrander+ TNR	MRS (wpm)	0.763	*P*= 0.000 (56± 29)	[113; -84]
Radner + TNR	MRS (wpm)	0.796	*P*= 0.000 (17± 23)	[27; -62]
MNread+ TNR	MRS (wpm)	0.751	*P*= 545 (3± 29)	[27; -59]
Bailey-Lovie+ TNR	MRS (wpm)	0.590	*P*= 0.000 (-84± 25)	[-35; -133]

Table (3.5) shows there is a statistically significant effect of the chart type on RA, MRS, CPS and logRAD/logMAR ratio. For measuring RA, statistically significant differences were found between all the charts, except between Colenbrander and MNread the difference was not statistically significant, but the correlation was poor (r = 0.292). The correlation coefficients for reading acuity ranged from poor to moderate correlation between the charts. Low Bland-Altman agreement for RA results was found between Colenbrander and Radner, Colenbrander and Bailey-Lovie and MNread and Bailey- Lovie. On the other hand, high agreements were found between Colenbrander and Bailey-Lovie, Radner and MNread and Radner and Bailey-Lovie (Figure 3.10).The Colenbrander chart usually resulted in better reading acuity compared to the other charts, while Bailey-Lovie resulted in worse reading acuity.

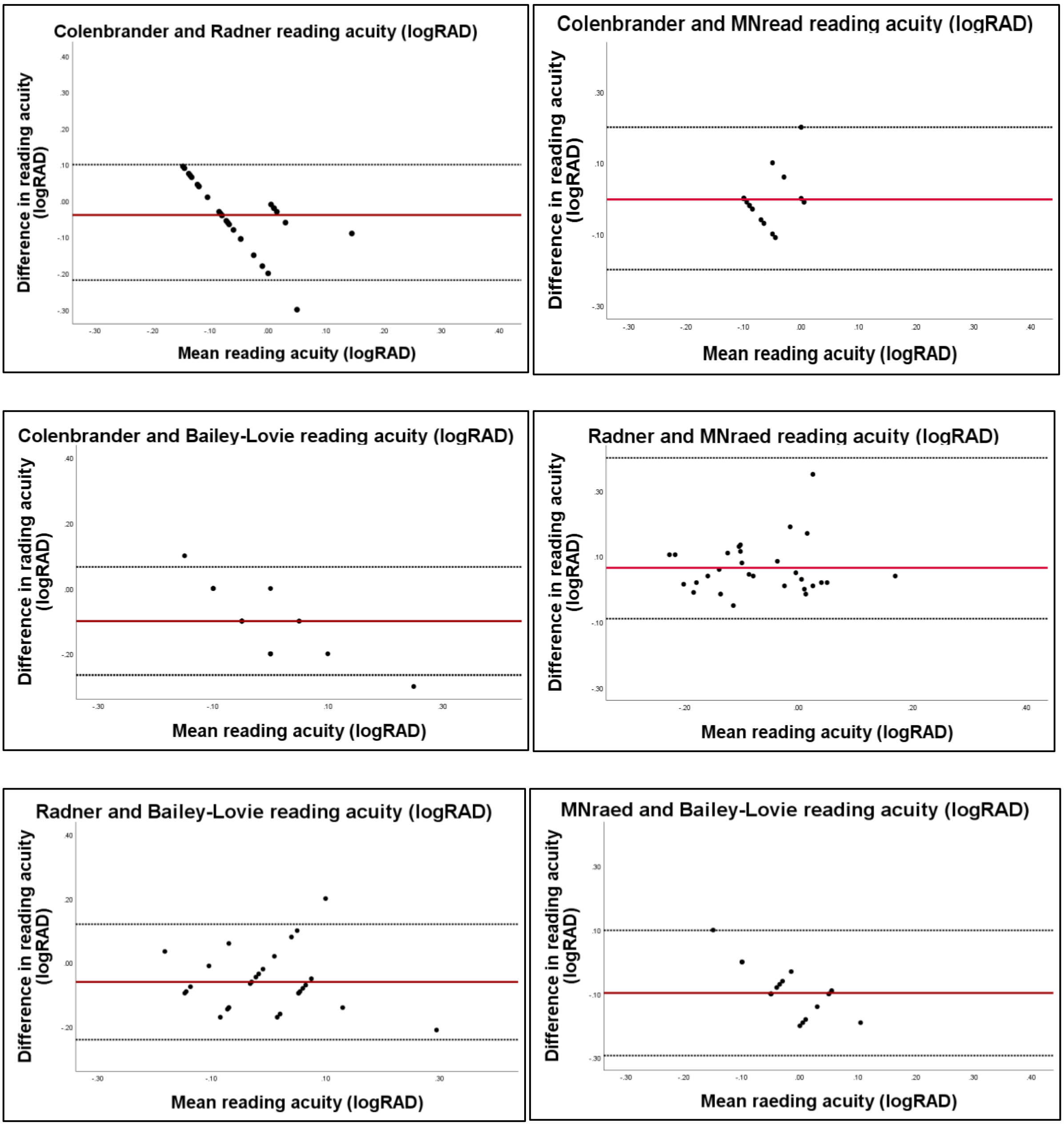


Figure 3.10: Bland-Altman agreement plots of the reading acuity between the charts. The differences of the mean between the charts is plotted against the mean reading acuity of both charts

Similarly, statistically significant differences between the charts were revealed for measuring MRS ($p<0.05$). Although high correlation and agreement were found between Colenbrander and Radner (r = 0.818) and Radner and MNread (r = 891), the difference in average maximum reading speed was statistically significant (p = 0.00). This difference was wpm: 37 ± 26, 95% confidence interval CI: (63; 83) and wpm: -14 ± 19, 95% CI: (-21; -7), respectively. In the Bailey-Lovie chart, the MRS result was significantly slower than continuous text reading charts (Figure 3.11). Poor correlation (r = 0.438) and agreement were found between Colenbrander and MNread. The difference was statistically significant wpm: 110 ± 40, 95% CI: (95; 126).

Colenbrander charts revealed significantly faster reading speed compared to the other charts (Figure 3.11). Bland-Altman plots (Figure 3.12) showed a wide data spread for MRS between TNR and the other charts, except for the comparison between Bailey-Lovie and TNR the agreement was high. However, statistical significant difference was found (p = 0.000).

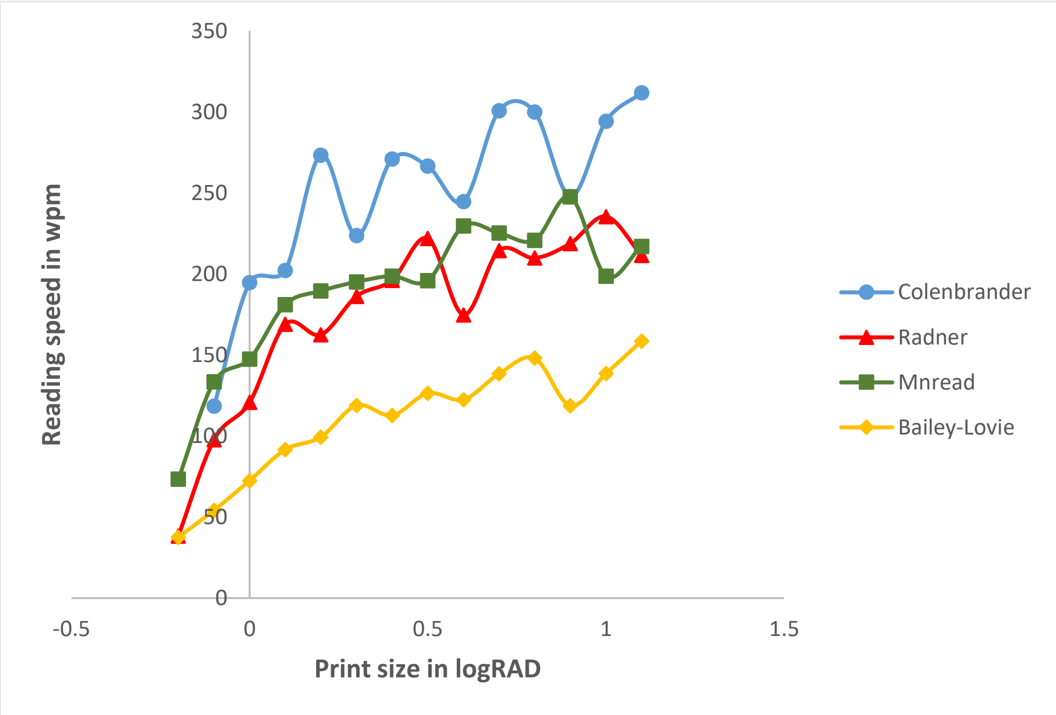
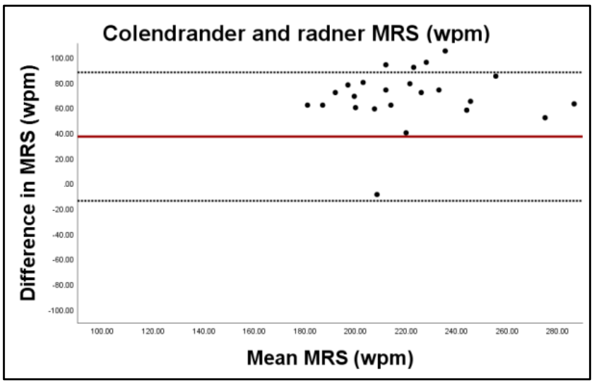
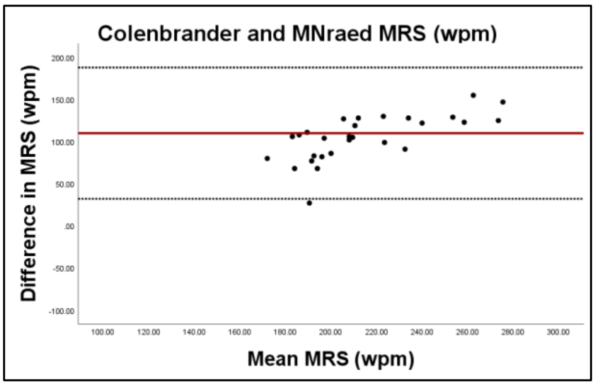


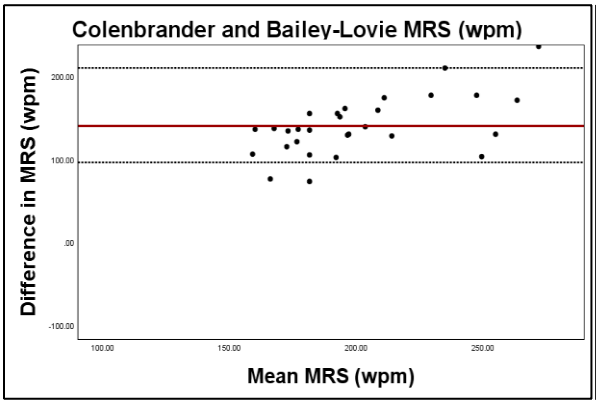
Figure 3.11: Mean reading speed in wpm for 29 normally sighted participants plotted as a function of print size for Colenbrander, Radner, MNread and Bailey-Lovie

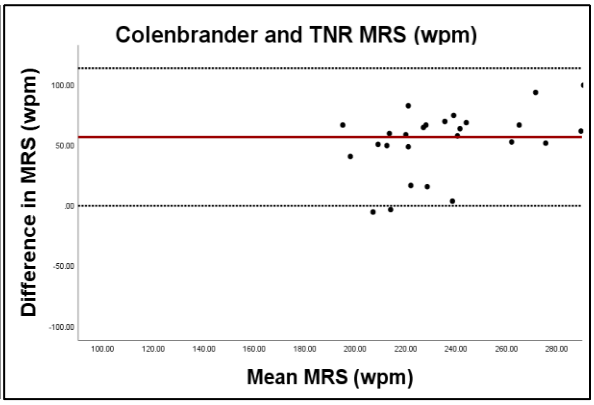
For measuring CPS, poor Pearson correlations coefficients were found between all the charts, ranging from r = 0.272 to r = 0.444. No significant differences were found for CPS between Colenbrander and Radner, Colenbrander and Bailey-Lovie, Colenbrander and MNread and Radner and MNread. However, the differences were significant between MNread and Bailey-Lovie and Radner and Bailey-Lovie. See the Appendix 2.1 for Bland-Altman agreement plots between the charts for measuring CPS.

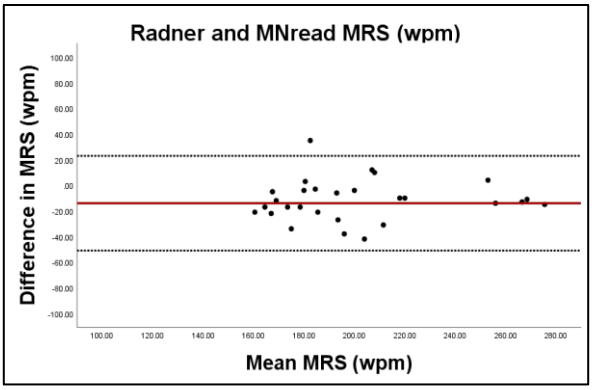
As shown in table (3.5), good Bland-Altman agreements were found for measuring logRAD/logMAR ratio between the reading charts. The correlations between the charts range from low (r = 0.324) to moderate correlation (r = 0.658) for logRAD/logMAR ratio results. No strong correlation was found. Statistical significant differences were found between Radner and Bailey-Lovie, Radner and MNread, MNread and Bailey-Lovie and Colenbrander and Bailey-Lovie. However, no significant differences were found between Colenbrander and Radner and Colenbrander and MNread. The logRAD/logMAR ratio for Bailey-Lovie charts significantly reported lower ratio compared to the other charts. See the Appendix 2.2 for Bland-Altman agreement plots between the charts for measuring logRAD/logMAR ratio.

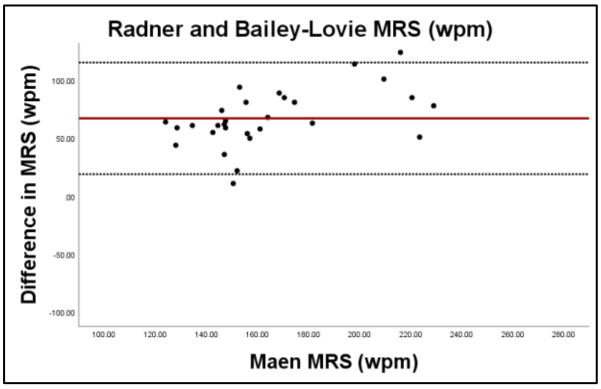












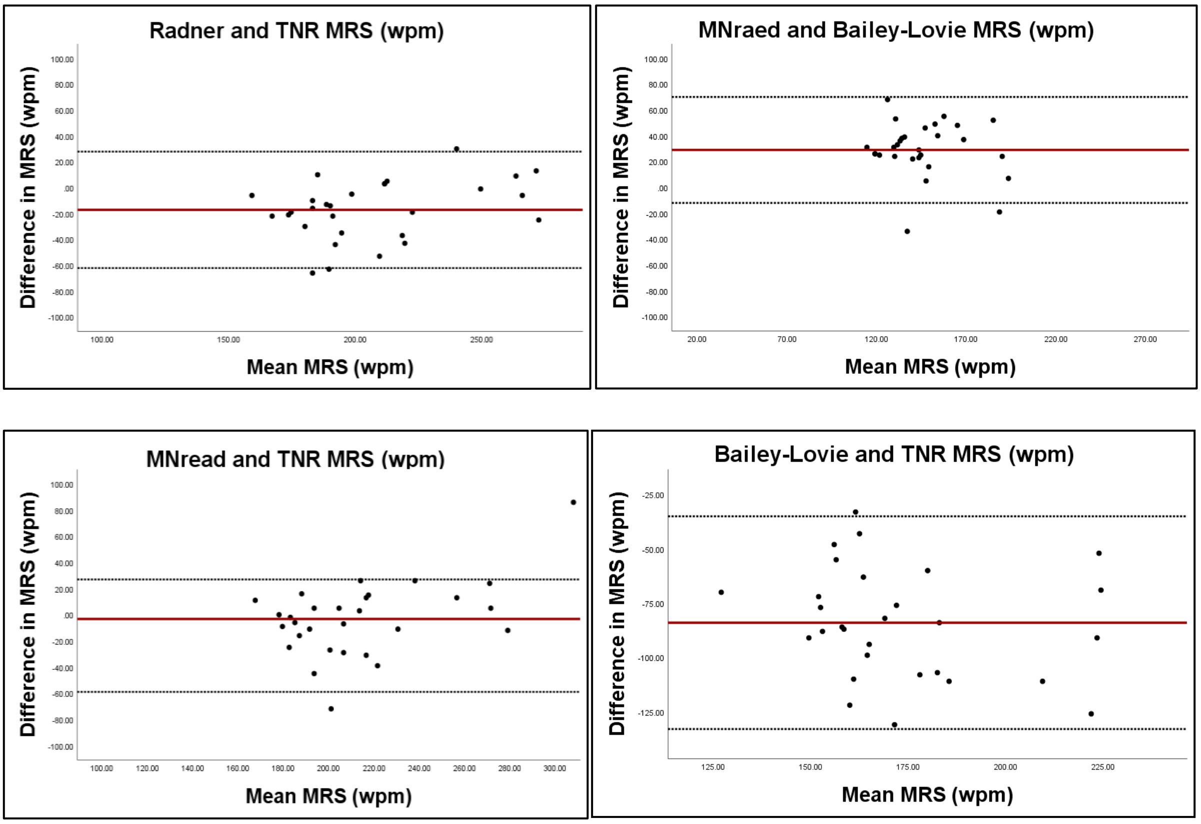


Figure 3.12: Bland-Altman agreement plots of the reading speed between the charts. The differences of the mean between the charts is plotted against the mean reading speed of both charts

3.2.5 Correlation of the mean reading speed between the short sentences charts and long paragraph chart (IReST)

To assess the validity of the reading speed results, this study compared the reading speed results obtained from a short sentences paragraph (Radner, MNread, Colenbrander and Bailey-Lovie) with the long paragraph reading test (IReST). A Times New Roman chart can be considered as a long paragraph test because the number of the words in each sentence ranges from 5 to 80 words. The correlation of reading speed results was also calculated between IReST and TNR. Because the only parameter that can be calculated from IReST is the average reading speed of the ten paragraphs, average reading speed parameter was compared between the charts.

The averages of the reading speed were normally distributed in IReST, TNR and Colenbrander (Shapiro-Wilks test p >0.05), whilst with the Radner, MNread and Bailey-Lovie they were not normally distributed (Shapiro-Wilks test p <0.05). Therefore, non-parametric Spearman correlation was used in the analysis. The correlation between IReST and the other charts was significantly high p <0.05. The correlation between IReST and Colenbrander was r_s = 0.765; IReST and MNread, r_s = 0.779; IReST and Radner, r_s = 0.864; IReST and Bailey-

Lovie, r_s = 0.762; and IReST and TNR, r_s = 0.764. Although, the correlation between IReST and Bailey-Lovie was high, Figure (3.16) shows that the distribution of reading speed of 29 participants was relatively widely spread between the two tests. Figure (3.14) and (3.15) show that the strongest correlations were for Radner and MNread.

A Wilcoxon signed-rank test was conducted to determine if the median differences of the reading speed between the long paragraph test (IReST) and short sentences tests are statistically significant. The 29 young normally sighted participants read the IReST significantly faster than MNread (median difference = 19 wpm, $p<0.05$), Radner (median difference= 32 wpm, $p<0.05$), and Bailey-Lovie (median difference = 90 wpm). This is opposite of what was found between IReST and Colenbrander. The participants read the IReST significantly slower than Colenbrander (median difference= 39 wpm) (Fig 3.13). No statistically significant difference was found between IReST and TNR (Fig 3.17).

Since the IReST reading chart designed to measure the reading speed in one print size which is the newspaper print size (0.4 logMAR), this study compared the reading speed of single sentences of the same print size in Colenbrander, Radner, MNread and Bailey-Lovie versus the IReST text paragraph. Figure 3.18 shows that the participants read the IReST paragraph significantly faster than 0.4 logMAR single sentence in MNread (median difference = 19 wpm, p= 0.011), Radner (median difference= 36 wpm, p= 0.003), and Bailey-Lovie (median difference = 100 wpm, p< 0.001). However, the participants read the IReST significantly slower than 0.4 logMAR single sentence in Colenbrander chart (median difference= 47 wpm, p< 0.001). This findings is consistent with that observed in the previous paragraph which compares the reading speed of IReST to the average reading speed of all print sizes in other reading charts.

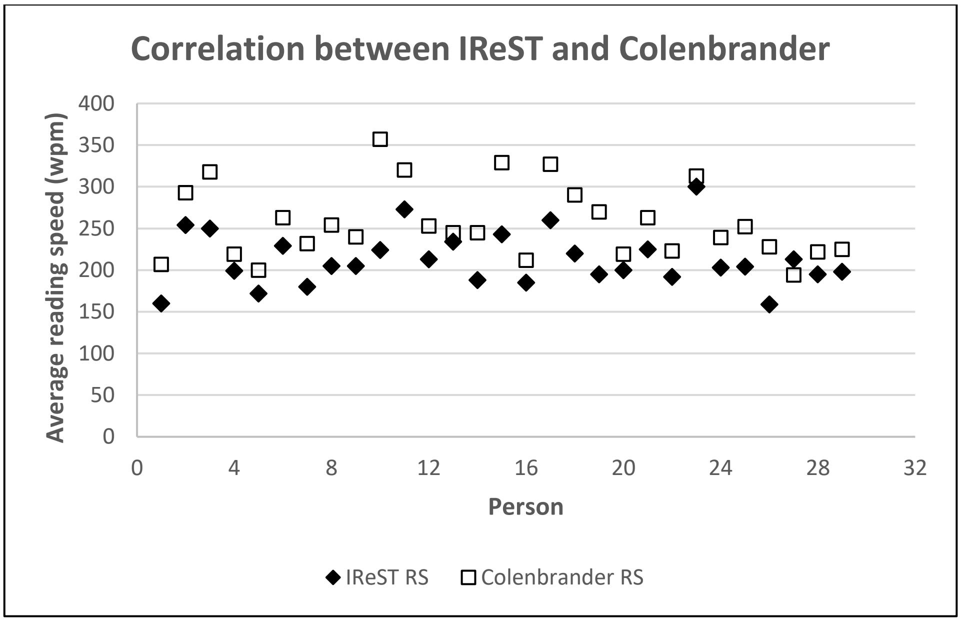


Figure 3.13: Reading speed distribution of the 29 participants. Long paragraph (IReST) vs. short sentences (Colenbrander)

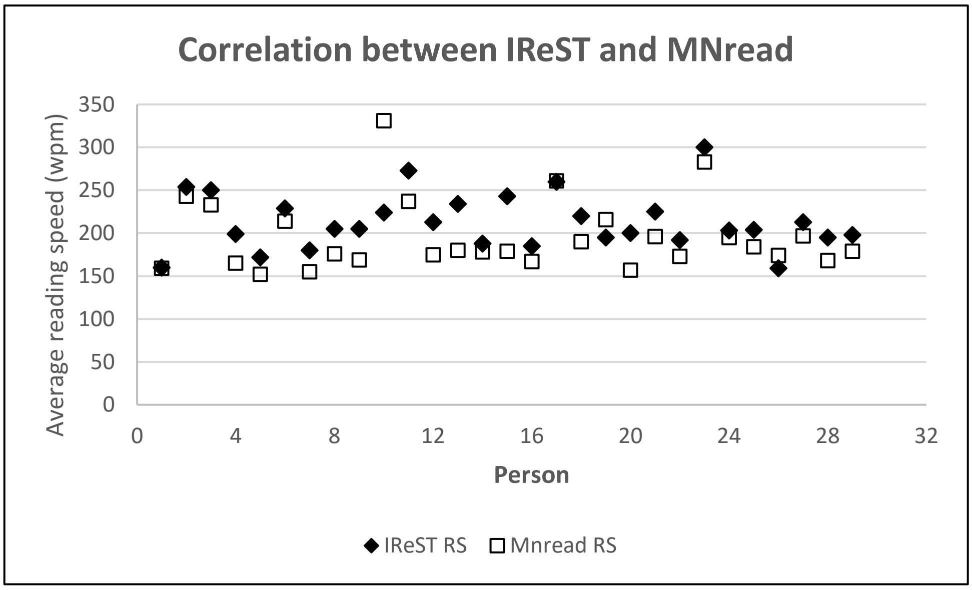


Figure 3.14: Reading speed distribution of the 29 participants. Long paragraph (IReST) vs. short sentences (MNread)

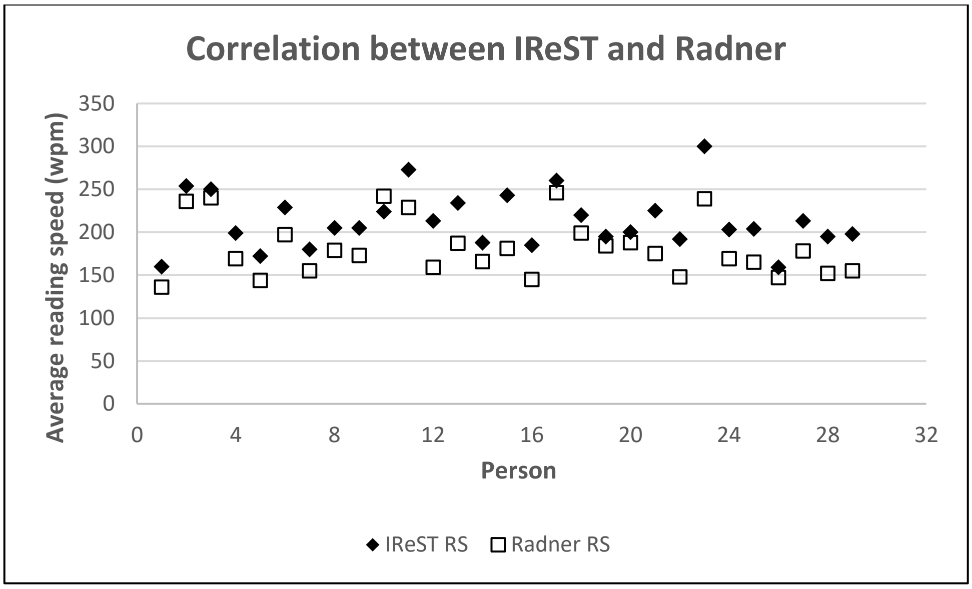


Figure 3.15: Reading speed distribution of the 29 participants. Long paragraph (IReST) vs. short sentences (Radner)

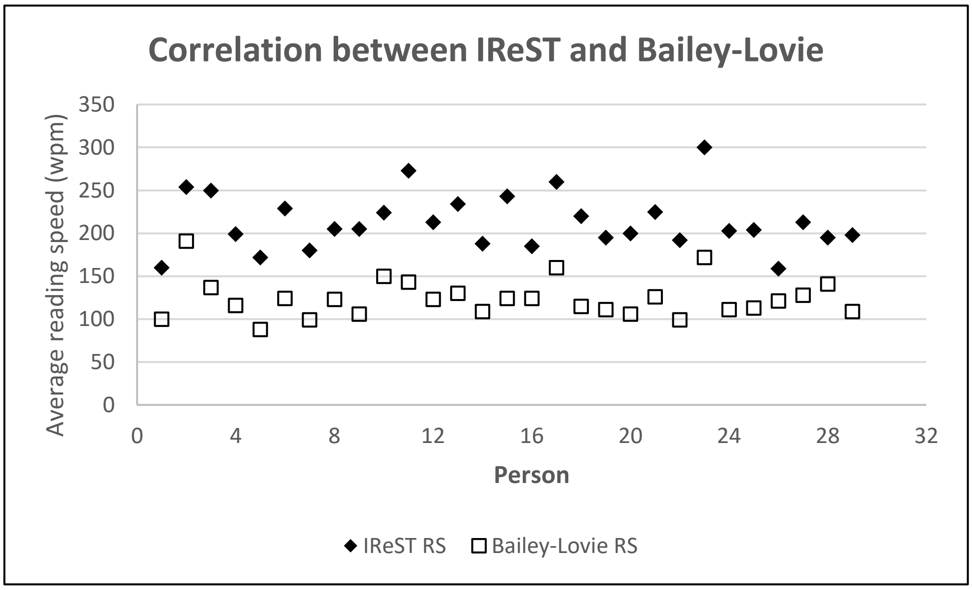


Figure 3.16: Reading speed distribution of the 29 participants. Long paragraph (IReST) vs. short sentences (Bailey-Lovie)

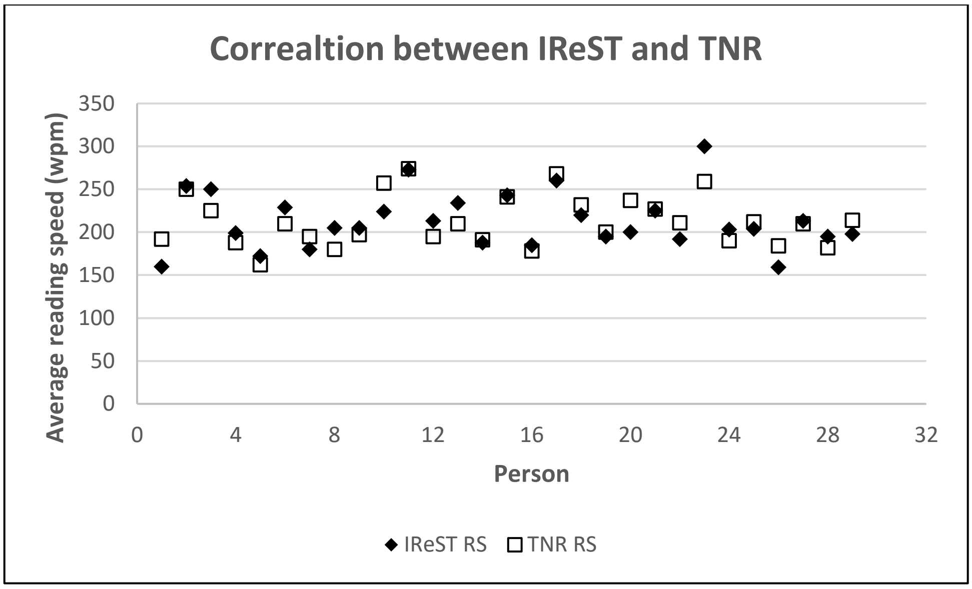


Figure 3.17: Reading speed distribution of the 29 participants. Long paragraph (IReST) vs. TNR

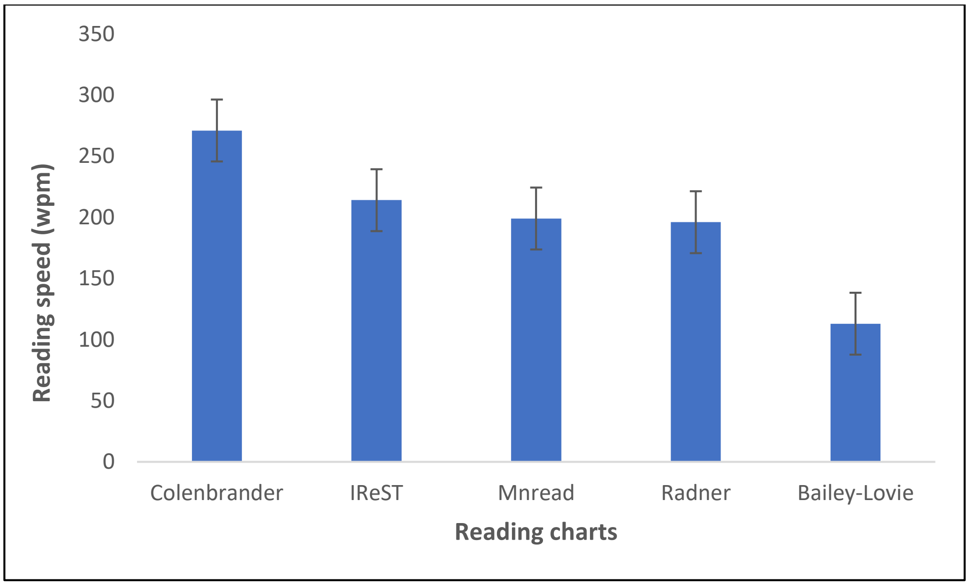


Figure 3.18: Reading speed of the IReST reading chart compared to 0.4 logMAR single sentence in MNread, Radner, Colenbrander and Bailey-Lovie

3.2.6 Comparison of each IReST text paragraph repeatability and reliability

Statistical analysis was performed with respect to the normally distributed data of the reading speed of each text in IReST.

The results of the repeatability analysis of the IReST 10 paragraphs are shown in Table (3.6). The test-retest correlations for measuring reading speed for all the paragraphs were high ($r \geq 0.7$) except for three paragraphs (text No.5, r = 0.515; text No.6 r = 0.686; text No.10 r = 0.683). From text number 1 to text number 5, statistically significant differences were found between the two testing session $p < 0.05$. The differences between the two testing session in the first five texts in the IReST are highlighted in Figure (3.19). The Bland-Altman agreement was low in text No. 1, text No.2 and text No.4 (Fig 3.20). Low correlation and Bland-Altman agreement with statistically significant difference was found between the two testing sessions in text number 5. From the table it can be seen that from text number 6 to text number 10, resulted in high agreements without any statistically significant differences between the two testing sessions for measuring reading speed. In general, the reading speed results in the second session were faster than the first session. The Bland-Altman plots of the remaining text paragraphs are given in the Appendix 2.3

The reliability analysis of the ten text paragraphs resulted in a good Cronbach's alpha ranging from 0.935 to 0.807 except for text number 5 where the reliability was lowest (α = 0.675).

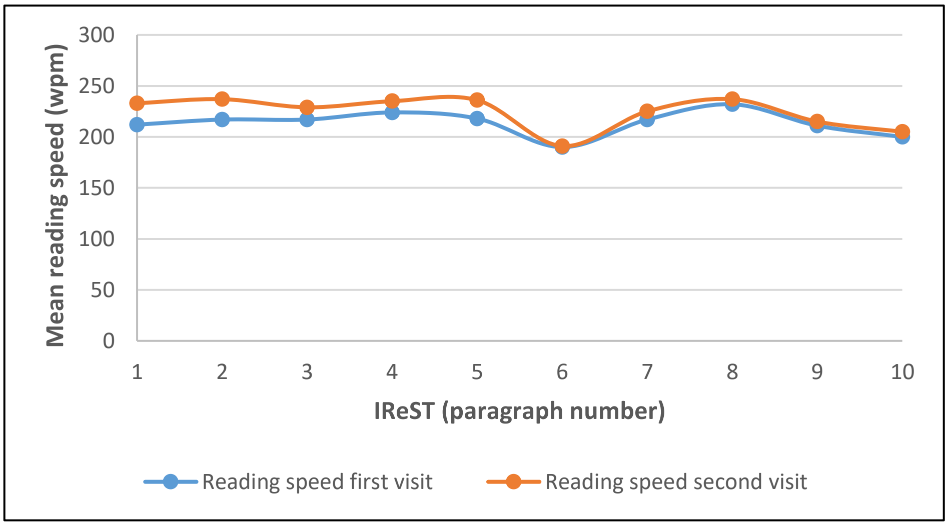


Figure 3.19: Mean reading speed of each text in IReST reading test in the two testing sessions

Table 3.6: Repeatability of reading speed results of each text in the IReST reading chart. Pearson correlation coefficient and Bland-Altman agreement were undertaken between the two testing sessions. The differences of the reading speed measurements between the two testing sessions were analyzed for significance by means of t-test. The cut-off level for statistical significance was p <0.05, one-tailed t test (second visit- first visit).

Number of text	Person correlation (r)	t test (mean±SD), p value	Limits of agreement
1	0.868	(20±19), p = 0.000	[57; -17]
2	0.821	(19±26), p = 0.000	[73; -25]
3	0.867	(12±20), p = 0.004	[35; -27]
4	0.892	(10±19), p =0.005	[47; -27]
5	0.515	(18±46), p = 0.049	[108; -72]
6	0.686	(1±25), p = 0.753	[50; -48]
7	0.685	(8±33), p = 0.204	[72; -56]
8	0.733	(4±29), p = 0.419	[60; -52]
9	0.714	(3±29), p = 0.445	[55; -50]
10	0.683	(5±29), p = 0.350	[61; -51]

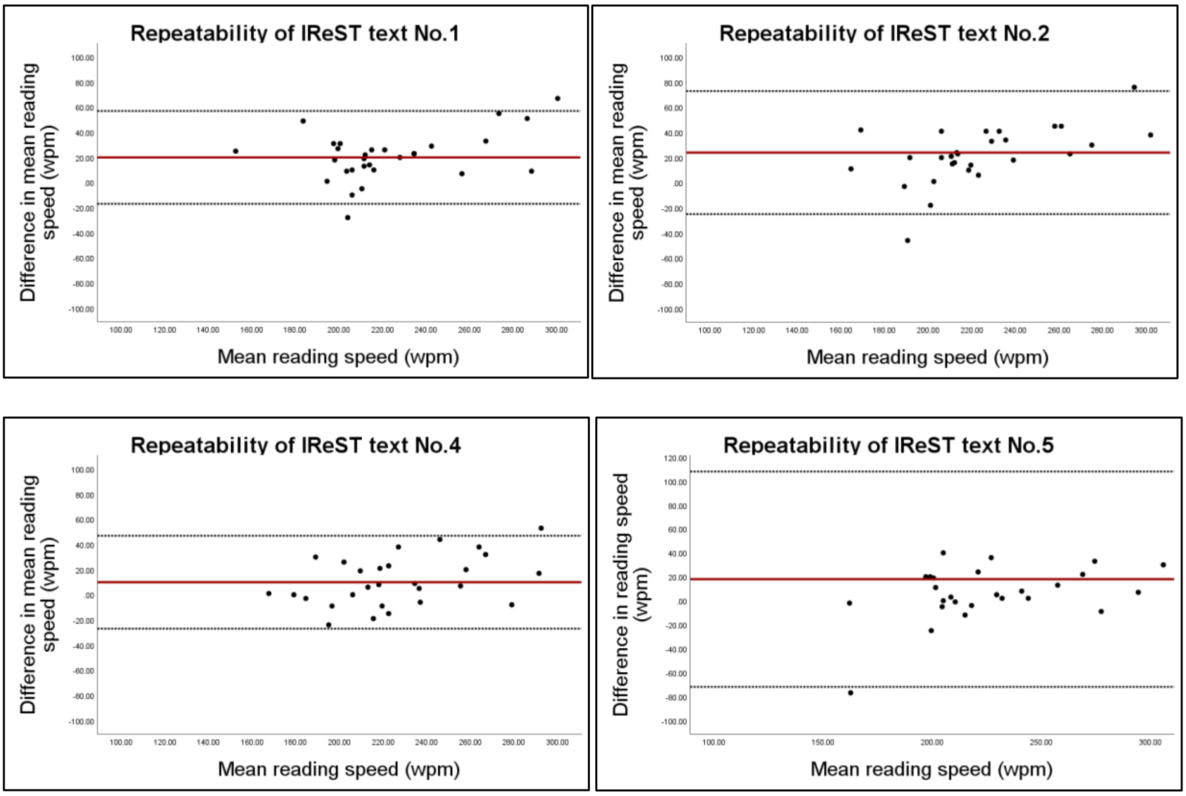


Figure 3.20: Bland-Altman agreement shows a low agreement between the two testing session in IReST text No. 1, 2 and 4 and 5.

3.3 Discussion

3.3.1 Repeatability and reliability of the reading charts for measuring the reading performance metrics

The use of a standardized reading test for measuring the reading performance metrics is of significant scientific and clinical value. An overview of the history and development of reading tests has been well established and described by Rubin (2013). His review was in agreement with Brussee et al. (2014) and they confirmed that studies describing the reliability and repeatability of the available reading tests are lacking.

This study assessed the test-retest repeatability and inter-chart reliability in the reading test charts for all reading performance metrics in normally sighted participants by using same study design and same algorithm for scoring reading performance metrics in each chart. This study found that the test-retest and the inter-chart reliability for different reading tests were varied, with most of the variance being attributed to the charts themselves and not to inter-individual variability. In our point of view, evaluation of the reading test charts should be first calibrated among the normally sighted people to eliminate any factors that may affect the reading performance measurements. To standardize the reading tests, the print size should be the only parameter affecting the reading performance measurements throughout the chart.

The present study is the first study to evaluate the inter-chart and test-retest reliability of a reading charts (in English language) under the same clinical conditions for all reading performance metrics.

For the Radner chart it should be noted that although the correlations between the two testing sessions for measuring RA and MRS were high, a low Bland-Altman agreement was found with statistical significant differences between the measurements. These findings are contrary to that of Stifter et al. (2004a) who found a high test-retest reliability for all reading performance metrics. However, they assessed the Radner chart in (German language) and on a low vision subjects. It is well known that different languages have a different word length. Another possible explanation of that higher repeatability is that on all Radner reading charts there has been a sentence repetition which may reduce the variance and lead to higher repeatability (Maaijwee et al., 2008a). In this study the reliability of Radner chart was high, and this is in agreement with previous studies (Radner et al., 2002, Maaijwee et al., 2007, Alió et al., 2008).

It is interestingly that the CPS showed a lower repeatability and reliability in all reading charts. This is in agreement with other studies in which the largest part of the variability in CPS measurements was explained by unidentified sources (Stifter et al., 2004a, Subramanian and Pardhan, 2006, Mataftsi et al., 2013). One explanation is the examiner’s subjective decision as CPS has to be set as smallest print size the participants can read with optimal reading speed.

MNread chart showed a very high reliability and repeatability for measuring RA, MRS and LogRAD/LogMAR ratio without any significant differences between the two testing sessions. As mentioned earlier lower repeatability result for CPS was found in MNread chart. However, the study of Subramanian and Pardhan (2006) showed a better repeatability result for measuring CPS and this might be explained by analysis of repeatability was evaluated by testing twice on the same day. It has been recommended that the minimal interval between test-retest should be about 10 days to eliminate learning effect (McMonnies, 2001).

Although the inter-chart reliability and repeatability of the Colenbrander reading charts have not been investigated, a number of studies have developed new continuous text reading charts in different languages based on the Colenbrander chart criteria. Examples include the study by Abdulkader who developed a standardized Arabic reading chart with a layout similar to that of the Colenbrander chart (Alabdulkader and Leat, 2017). This study found lower correlations between the two testing sessions in measuring RA and MRS compared to MNread and Bailey-Lovie chart and Radner. For CPS all the reading test scored low in repeatability but in Colenbrander no correlation was found between the two testing sessions.

The purpose and design of the IReST chart is different in comparison with other reading test charts. The repeatability for IReST has not yet investigated in the previous studies. This study found very high repeatability (r= 0.918) and also reliability (α= 0.951) for reading speed measurements for IReST. This study analyzed the outcomes for the average of reading of all the 10 long paragraphs in IReST which may be infeasible to apply in clinical condition, a subsequent section of this study will assess the repeatability and reliability of each text in IReST to obtain results that are comparable and not affected by choice of the text paragraph.

Bailey and Lovie (1980) believed that evaluating the reading performance with unrelated words is better than meaningful sentences, because the different individual reading behavior, which is not related to reading ability and the contextual cues, assists the guessing during a reading test and, thus, could lead to overestimating of the near acuity. As mentioned in Chapter 1, no studies were found in the literature that evaluate the Bailey-Lovie unrelated words reading chart characteristics in term of reliability and repeatability. This study found that a very high reliability for Bailey-Lovie chart for measuring reading performance metrics even for CPS (α= 0.717). However, significant difference between the two testing sessions was found for RA and LogRAD/LogMAR ratio. Regarding the RA this difference is not clinically relevant, as they were below one line of reading acuity 0.1 log unit.

Although the Times New Roman reading chart has not been developed in accordance with the recommendations for standardized reading charts (Colenbrander and Committee, 1988), it was included in this study based on our previous findings which showed frequent use of this type of chart by optometrists in the UK. The results revealed that Times New Roman chart is not a good choice for measuring the reading performance.

3.3.2 Comparison of the reading errors between the charts

In this comparison, we relied on the results of reading errors made in the first session rather than the second session to eliminate any possibility of the practice effect. The most reading errors were made with Radner and the least reading errors were made with the Bailey-Lovie unrelated word reading chart with exclusion of the IReST, as the reading errors were counted for the all ten long paragraphs which resulted in more reading errors due to more text being read. This result is contrary to that of Brussee et al. (2015) who found that the Radner chart had the lowest reading mistakes compared to the Colenbrander and IReST. However, their study was in the Dutch language and they counted the reading errors as mistakes per 100 characters. The present study counted the mean of the reading errors that were made on whole chart and this method is more practical and may have a more clinical importance.

As mentioned earlier, the aspect of the mechanics of running Bailey-Lovie unrelated word reading chart make the test rather more difficult than continuous text reading charts. The combination of the unrelated words slows down the reading speed and makes the participant concentrate more on the upcoming unexpected word while reading the sentence and this may explain why the lowest number of reading errors was made with the Bailey-Lovie chart. Whereas, in the continuous text reading charts, because of the uniform format of the meaningful sentences, it may accelerate reading speed and result in more errors. It has been found that, during the reading test, the ability to use prediction of the next word determines the reading speed and differentiates between the non-fluent and fluent reader (MacKeben et al., 2015). The influence of the reading material difficulty on the reading errors can be excluded, because all the reading charts were at the ability level of the young normally sighted participants.

3.3.3 Comparison of reading performance outcomes between reading test charts

One aim of this study was to investigate the agreements between the five reading test charts in normally sighted participants. For all the reading tests, the agreements of the reading performance metrics were varied between the reading tests.

The results showed that the Bailey-Lovie reading charts resulted in lower reading acuity and slower reading speed compared to the other charts (Fig 3.11). This result is in the line with previous study obtained by Mackeben et al. (2015) who also found that the reading speed for unrelated word reading chart was always slower than continuous text. There are two possible explanations of this result. First, the contextual cues play an important role to encourage the guessing, especially for fluent readers as all the participants in this study were university students. Bailey and Lovie believed that reading unrelated words charts reduces the inter-subject variation caused by reading skill level or educational status, which are not related to the visual abilities. Therefore, the continuous text reading charts may lead to faster reading speed and overestimate the reading acuity. However, the continuous text reading charts are more representative for daily reading material compared to the Bailey-Lovie unrelated word reading charts. Second, as mentioned earlier in the literature review the Bailey-Lovie charts are criticised in terms of the word length (more than 10 letters) and being difficult to read (Rubin, 2103) which may result in slower reading speed.

One interesting finding was that the Colenbrander reading charts always scored significantly higher maximum reading speed compared to the other reading charts (Fig 3.11). A possible explanation for this finding might be that the Colenbrander reading chart makes use of relatively short words compared to the other reading charts (Brussee et al., 2015).

In the previous studies, there are no direct comparisons available between MNread and Radner charts. The results revealed that the Radner chart resulted in significantly slower MRS than MNread (p= 0.001). This difference was wpm: -14 ± 19, 95% CI: (-21; -7). Although the Radner chart used a sans-serif typeface and MNread used a serif typeface, it was reported that the typeface has little or inconclusive influence on reading speed (Rubin et al., 2006). During the experiment, it was noticeable that the majority of the participants were confused and slowed down the reading when they read this sentence in the Radner reading chart 1: ‘our <u>sister</u> travelled here in <u>his</u> helicopter, for which <u>he</u> bought leather seats recently’. The participants always made a reading mistake in this sentence and said ‘her’ instead of ‘his’ and ‘she’ instead of ‘he’ as opposed to the word ‘sister’. Another noticeable thing in Radner reading chart 3 is that the sentence ‘big animal hunted near the old rainforest, in which we found many impressive trees‘ is repeated two times in the same chart and this could raise the effect of learning practice on reading speed. However, for measuring reading performance, the print size should be the only parameter that affects reading outcomes throughout the chart. This means that, when the reading chart is presented to the patient, all the sentences should be equally difficult to read. This also accords with Brussee et al.’s (2015) study which showed that, in the Radner reading chart, there was a significant deviation from the sentences’ means. They reported that two sentences were read significantly slower and four sentences were read significantly faster in the Radner chart, which means the sentences were not equally difficult to read within the test.

Although the data of this study confirmed poor reliability and repeatability of TNR chart, the findings of the low agreement for measuring MRS between the TNR and other reading charts must be interpreted with caution because TNR consists of paragraphs while MNread, Radner, Colenbrander and Bailey-Lovie consist of short sentences. This fact makes the comparison between paragraphs and sentences not reasonable. Brussee et al. (2015) reported that it is important to determine the clinically relevant differences in terms of reading speed between paragraphs and sentences to give insight into the method of scoring reading performance metrics. Therefore, it could not be argued that the differences and low agreements between

TNR and other reading charts for measuring MRS were due to the poor characteristics of the TNR chart.

Measurement of the reading acuity is important to compare treatment outcomes in clinical practice or researches. However, the comparison between the reading charts for measuring reading acuity demonstrated statistically significant differences ($p<0.05$), except between the Colenbrander and MNread where the agreement was high without significant differences. These results are likely to be related to the design of the reading charts in terms of the smallest print size. The smallest print sizes are varied between the charts with the ending print sizes being Colenbrander (-0.1 logMAR), Radner (-0.2 logMAR), MNread (-0.5 logMAR) and for Bailey-Lovie (-0.2 logMAR at 40cm). Because the participants of this study were young, healthy and fluent readers, it was expected to have good near visual ability. Therefore, for example, a participant who was able to read the smallest print size in Colenbrander chart (-0.1 logMAR) may have the ability to read smaller print sizes. This fact may explain the difference in reading acuity measurement between the charts. Despite this result, a question remains as to whether the agreement results of reading acuity between the reading charts will be similar for a presbyopic group of subjects. This will be investigated in the next chapter. According to this variation in near visual acuities between the charts, logRAD/logMAR ratio also showed a significant difference between the reading charts.

A clear finding was the weak correlation between the reading charts for measuring CPS. This result is likely to be related to the scoring rule method of CPS as the CPS is not a direct measurement, such as reading speed or reading acuity, since it has to be determined from the graphical plot of the reading speed against the print sizes as the smallest print size that can be read with maximum reading speed. However, in Figure (3.11) Radner, MNread, Colenbrander and Bailey-Lovie show a decrease in reading speed with decrease of print size and this is a good indicator that the slowing of reading speed results are likely to be due the change in print size rather than chart characteristics (Radner et al., 2002).

This result provides evidence for the research question as to whether the equality of the reading performance results depends on the chart used. The correlation and the agreements analysis of the reading performance metrics confirmed that reading chart outcomes need to be comparable with each other and different reading charts provide a different reading performance. Reading performance metrics measure the effectiveness of vision rehabilitation and therapeutic intervention (Rubin, 2013). Therefore, the best way to evaluate reading performance is using the same type of reading chart to facilitate the comparison of treatment outcomes. These findings confirm the need to homologate and standardize reading charts to provide a valuable tool in analyzing the researches concerning reading performance and the visual function in many fields of optometry.

3.3.4 Correlation of the mean reading speed between the short sentences charts and long paragraph chart (IReST)

In the present study, we analyzed the validity of the reading speed results obtained using short sentences test compared to long paragraph test in young healthy participants, and the results found high correlation. This high correlation is one of the most important criteria when developing a reading chart. Because, if the reading speed of the short sentences reading chart correlated well with the long paragraph test, such as IReST, it indicates a high validity of the test. The high validity of a standardized long paragraph reading test comes from the idea that, first, it represents a real life reading performance, including the navigation issues and fatigue effects, and, second, reading for a longer period of time decreases the variability (Trauzettel-Klosinski et al., 2012). This strong correlation indicates that the validity of the reading charts that have been chosen to be included in this thesis is high. However, the characteristics of these charts in terms of reliability and repeatability were different for measuring the reading performance metrics.

The reading speed distribution figures showed that the strongest correlation was between IReST and Radner, MNread, and TNR. The least correlation was found between IReST and Bailey-Lovie. The difference in reading speed between IReST and Bailey-Lovie was 90 wpm, and this is again because the Bailey-Lovie reading chart consists of unrelated random words rather than continuous text.

The results showed that the reading speed of the long paragraph test is faster than short sentences test charts. These results are in agreement with those studies obtained by (Munch et al., 2016) of the short text in the Danish language and by (Radner and Diendorfer, 2014) of short text in the English language. This outcome is contrary to that of Radner et al. (2002) who found the long test paragraphs were read slower than the short sentence reading test. The paragraphs in IReST test were extracted from reading book materials, which represents an integrated story. Therefore, IReST paragraphs might have higher contextual cues that result in increased reading speed in comparison with other charts.

Although the IReST long paragraphs test provides more information that is closer to everyday reading and has a high reliability and repeatability in measuring reading speed, the use of short sentence tests, such as Radner and MNread, is more convenient in research and clinical practice as, in addition to the reading speed, the reading acuity and critical print size can be obtained relatively quickly and provide more information about the visual function.

3.3.5 Comparison of each IReST text paragraph repeatability and reliability

The IReST reading test contains 10 unique paragraphs in one print size, and, in clinical practice or research, it may be not feasible to calculate the average reading speed from all the ten paragraphs, because this will consume more time. The practitioner may randomly choose certain paragraphs to assess the reading performance. This means that all the paragraphs should be equal in terms of reliability and repeatability. The paragraphs of the IReST reading test were evaluated in this regard.

IReST reading test was performed in 29 young normally sighted participants. The participants were asked to read the all IReST paragraphs in the two testing sessions, which were two to three weeks apart. The standardization of the selection of paragraphs of the reading test should be calibrated first among normally sighted people to eliminate any other factor that could affect the reading performance outcomes (Maaijwee et al., 2007).

The internal consistency and the validity of the IReST paragraphs were investigated and well-described in previous studies (Hahn et al., 2006, Trauzettel-Klosinski et al., 2012). However, the repeatability of each paragraph has not been described. Therefore, one aim of this study was to investigate the IReST text paragraphs repeatability in addition to the reliability (internal consistency).

This is the first study that has evaluated the repeatability of each paragraph in the IReST test. The results showed that the repeatability IReST texts are not equal. This leads to consideration of choosing specific paragraphs on the IReST reading test to measure the reading speed in research or clinical practice. The last five paragraphs in the IReST yielded better repeatability results than the first five paragraphs. Because it is difficult to figure out why certain paragraphs are different from the others in term of repeatability, the repeatability testing of each paragraph or sentence within the reading test is recommended to standardize the reading test. IReST paragraphs had a high equality in terms of reliability and this finding was also reported by other studies (Trauzettel-Klosinski et al., 2012, Brussee et al., 2015).

3.4 Conclusion

The present study confirmed that the clinical results of the reading performance metrics obtained with different reading charts are affected by the choice of the chart. The test-retest reliability for measuring reading performance metrics are higher in MNread chart, while the Bailey-Lovie word chart has a higher inter-chart reliability. For measuring reading speed only, the IReST chart provides very high repeatability and reliability results. However, repeatable text paragraphs must be chosen. This finding has important implications for evaluating the reading performance in clinical practice or scientific researches. The findings of the poor agreement between the reading performance metrics obtained during the same test session indicate that choosing the same test chart for consecutive examinations of the same patient or for comparison in research is very important in order to minimize the variation of the results.

Chapter 4

Reliability and repeatability of the reading performance metrics in different reading charts with presbyopic subjects

4 Introduction

Presbyopia is an age-related condition dependent upon the ability of the ocular lens to accommodate. Presbyopia causes inability to focus on the near objects due to the loss of the ocular lens flexibility. It especially decreases the reading ability, and one of the most important concerns of presbyopic sufferers is to return this to a comfortable level.

Recently, (Fricke et al., 2018) found that the global prevalence of presbyopia is one quarter of the world's population. The lifespan of our population is getting longer, and this increase highlights the importance of understanding lifestyle and environmental factors of this ageing population to ensure a high quality-of-life. The visual abilities of the older population have been well established by the Salisbury Eye Evaluation Project (Friedman et al., 1999, Rubin et al., 1997, West et al., 1997). Rubin et al. (1999) reported that visual functions including contrast sensitivity, visual acuity, visual field and glare, decrease with age. Friedman et al. (1999) demonstrated a detailed insight about reading performance of the older population wherein reading speed was less than 80 wpm in 10.8% of 2,107 presbyopic subjects reporting minimal difficulty in reading a newspaper, and, in fact, this is the lower limit for sustained reading performance.

Reading is an essential skill and is involved in many aspects of an older population, such as reading prescription bottles, food labels, mail or keeping up with current news and events. Visual acuity is well-maintained with increased age and the visual function of the older population is underestimated (Haegerstrom-Portnoy et al., 1999). This is supported by (Lott et al., 2001a) who showed that reading performance significantly decreased with increasing age even for people with good visual acuity. Therefore, ways for enhancing and measuring reading performance should be explored.

In recent years, there has been an increasing interest in presbyopia correction. To evaluate the efficacy of the presbyopia treatment, a reliable and repeatable reading test should be used in order to achieve meaningful results about the efficacy of the presbyopia correction. However, that is not yet the case as many different reading test charts are available.

Recent evidence discussed in Chapter three, found that different reading performance metrics were obtained when using different reading charts, in addition to which the repeatability and reliability of the reading charts were varied with young, normally sighted subjects. Therefore, there must be agreement on what reading test should be used to evaluate reading performance. MNread reading test scored better repeatability results while the Baily-Lovie unrelated words reading test scored better reliability results compared to the other charts with young, normally sighted subjects.

As the literature review makes clear, no direct comparisons between the reading tests were made. Furthermore, comparison between the studies using different subjects' characteristics and testing condition is difficult. Most previous studies evaluated the reading charts' repeatability and reliability restricted to young, normally sighted people. Thus, an exact comparison between the reading test charts is needed for presbyopic subjects. This chapter seeks to obtain data which will assist in addressing this issue.

The study in this chapter compares reading test charts' characteristics in terms of test-retest reliability and inter-chart reliability with uniform study design in older population aiming to provide clinicians and researchers investigating presbyopia, accommodation and presbyopia surgery with a standard reading test.

4.1 Methods

The proposed reading test charts (Colenbrander, Radner, MNread, Bailey-Lovie, IReST and TNR) were evaluated by performing test-retest and inter-chart reliability procedure. Two testing sessions (two to three weeks apart) were performed in presbyopic participants. Subjects aged over 40 years old participated in this study. The required sample size was calculated as in the previous chapter with G power analyser and with allow 10% for drop out for repeat testing resulted in a requirement of 30 subjects. Ethical approval was obtained from Aston University Ethical Committee (see Appendix 3) in accordance with the Helsinki Declaration and all the subjects gave their consent to participate in the study. All testing procedure was performed monocularly with the dominant eye wearing the optimal refractive correction. After performing the refraction, the participants were given an optimal presbyopic correction matching the test distance of 40cm. For participants wearing near vision spectacles, it was important to ensure that the glasses' prescription was up to date and the near addition compensated for the testing distance of 40cm. The same spectacles were used in the two testing sessions. The inclusion criteria were native English speaker subjects, educational level was at least A-level (secondary school qualification), best corrected distance visual acuity ≤ 0.1 logMAR and healthy eye. Healthy eye data were obtained by asking each participant if they had any known eye disease, undergone any ocular surgery, have binocular vision problems or experience any reading difficulty (dyslexia). Slit-lamp examination was performed to ensure that no media opacities were present. Amplitude of accommodation was measured using the RAF rule to ensure that the accommodation level was normal to the participant's age and they did not have any accommodative lag.

The material, measurement procedure, statistical analysis and the methods of calculating the reading performance metrics were the same as in Chapter three (Fig 4.1). For measuring the average reading speed using the IReST reading test, which contains 10 long paragraphs, three text paragraphs were chosen in the current study to assess the reading speed in presbyopic participants to eliminate the factor of the fatigue effect in this older age group. These three paragraphs were paragraph number six, seven and eight and were chosen according to the previous results in Chapter Three that showed that the last five paragraphs in IReST revealed better repeatability results compared to the first five paragraphs.

Figure 4.1: Experiment setting procedure. A headrest for the forehead was used to ensure constant viewing distance of 40cm. The non-dominant eye was occluded (Baashen, 2020).

4.2 Results

Reading performance measurements of 29 presbyopic participants were included, after exclusion of four subjects. Two subjects had an early sign of cataract, one subject had been diagnosed with age-related macular degeneration and the fourth had binocular vision problems. The participants age ranged from 41 to 70 years old (mean, 61.5 ± 8.8) years, 11 males and 18 females. There were 15 with right eye dominance and 14 with left eye dominance. Distance visual acuities were (mean,-0.08± 0.08) logMAR and near visual acuities were (mean, -0.01 ± 0.08) logMAR. The mean value of the spherical equivalent of the refractive error was (-0.25 ± 1.00) D. The amplitude of accommodation for all the participants was within a normal range for their age (3.3 D ± 0.9 D). The educational levels of the participants were 68% had a bachelor degree, 21% had A-level, 8% had a master degree and 3% had a PhD. All the participants were fluent readers.

4.2.1 Repeatability

The data of the reading performance metrics (RA, MRS, CPS and logRAD/logMAR ratio) for each reading chart were normally distributed (Shapiro-Wilks test $p > 0.05$). The differences of the reading performance metrics measurements between the two testing sessions, as well as the number of reading errors, were analysed for significance by means of t-test. The cut-off level for statistical significance was $P < 0.05$, two-tailed t-test. Pearson correlation coefficient and Bland-Altman agreement were performed to assess the repeatability of the reading performance measurements between the two testing sessions. Also, the coefficient of repeatability (CR) was calculated as defined by Bland and Altman (Bland and Altman, 1996) as 1.96 X $\sqrt{2}$ X standard error of mean (SEM). The differences in reading performance metrics between the two testing sessions is expected to lie below the CR for 95% probability.

4.2.1.1 Repeatability of continuous-text reading tests (Radner, MNread and Colenbrander)

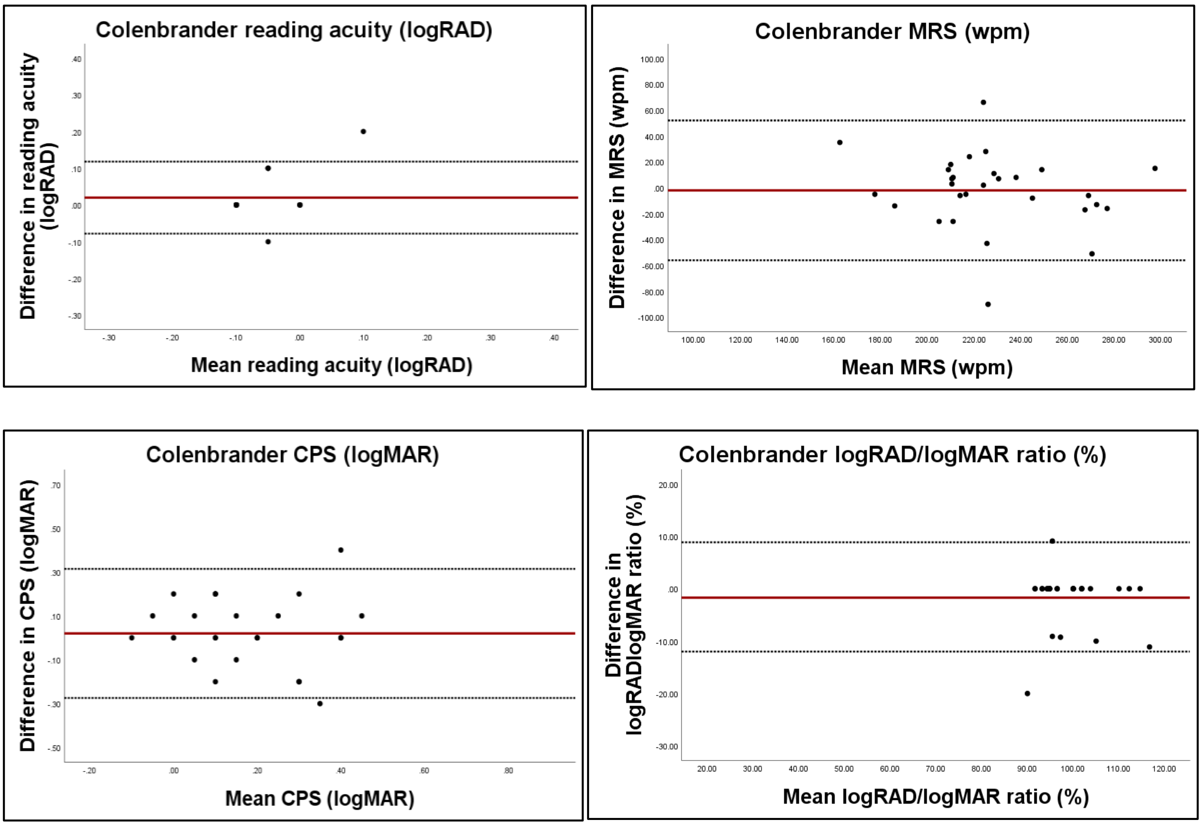
Table 4.1: Test-retest reliability between the first and the second testing session in presbyopic subjects. Pearson correlation coefficients n = 29

Reading metrics	Colenbrander (*r* value, *P* value)	Radner (*r* value, *P* value)	MNread (*r* value, *P* value)
Reading acuity (logRAD)	*r* = 0.620, *P* <0.001	*r* = 0.805, *P* <0.001	*r* = 0.845, *P* <0.001
Maximum reading speed (wpm)	*r* = 0.560, *P* <0.001	*r* = 0.781, *P* <0.001	*r* = 0.767, *P* <0.001
Critical print size (logRAD)	*r* = 0.599, *P* = 0.001	*r* = 0.638, *P* <0.001	*r* = 0.551, *P* = 0.002
LogRAD/LogMAR ratio (%)	*r* = 0.746, *P* <0.001	*r* = 0.807, *P* <0.001	*r* = 0.420, *P* = 0.023

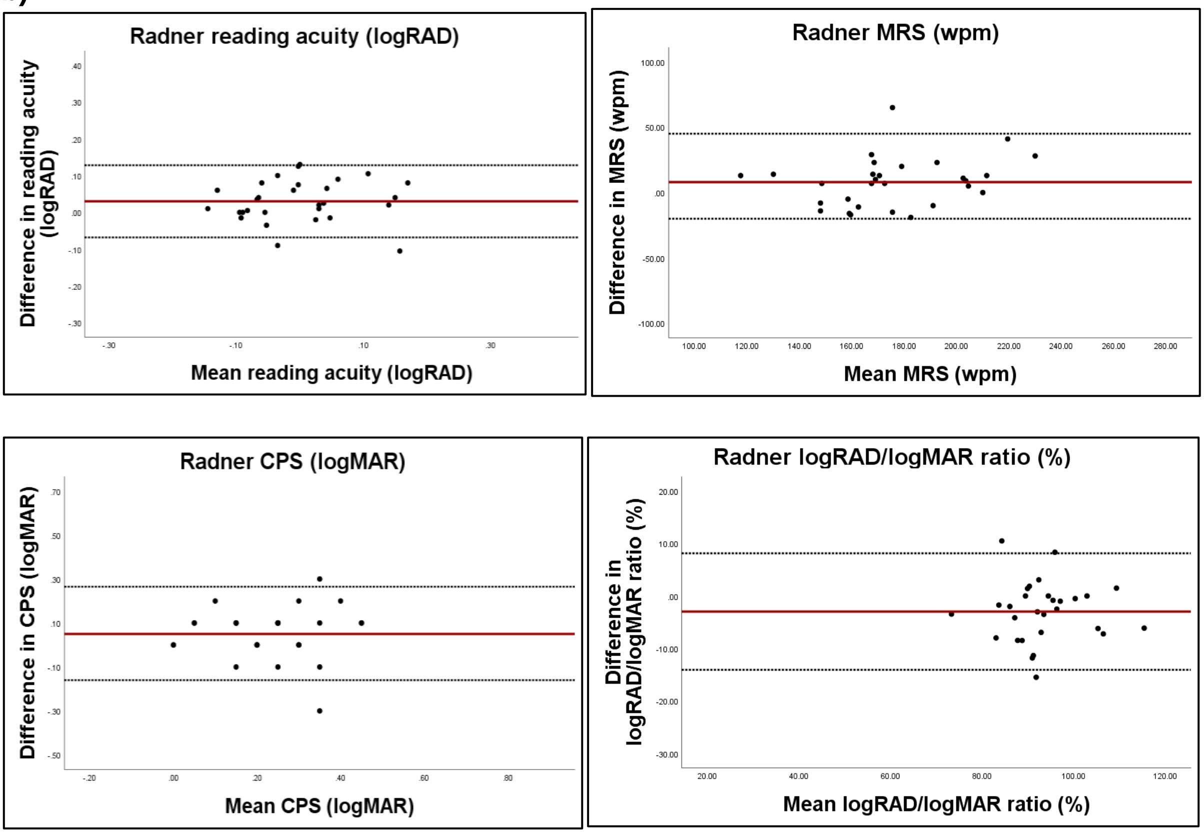
It can be seen from the data in the table above that, for measuring reading acuity, strong correlation was found between the two testing sessions in Radner (*r* = 0.805) and MNread (r = 0.845) while moderate correlation was found in Colenbrander (*r* = 0.620). For maximum reading speed measurements, the correlation in the Colenbrander (*r* = 0.560) was lower than Radner (*r* = 0.781) and MNread (*r* = 0.767). The test-retest correlations for the three reading test charts were moderate for measuring CPS. For logRAD/logMAR ratio, very weak correlation was found between the sessions in MNread chart (*r* = 0.420) while high correlations were found in Radner (*r* = 0.807) followed by Colenbrander (*r* = 0.746). All the correlations were tested for statistical significance *P* value <0.05, as shown in Table 4.1.

No statistically significant differences between the two sessions were found in Colenbrander and MNread for measuring all the reading performance metrics. However, in the Radner reading chart, statistically significant differences were found in all the reading performance metrics. These differences were in RA LogRAD: 0.03 ± 0.09, 95% confidence interval CI: (0.01; 0.05) *P* = 0.007, in MRS wpm: 8 ± 19, 95% confidence interval CI: (1; 15) *P* = 0.028, in CPS LogMAR: 0.05 ± 0.11, 95% confidence interval CI: (0.008; 0.09) *P* = 0.023 and in LogRAD/logMAR ratio (%): 2.79 ± 5.6, 95% confidence interval CI: (-5.13; -0.82) *P* = 0.009.

a)


Colenbrander reading acuity (logRAD)
Difference in reading acuity (logRAD)
Mean reading acuity (logRAD)
Colenbrander MRS (wpm)
Difference in MRS (wpm)
Mean MRS (wpm)
Colenbrander CPS (logMAR)
Difference in CPS (logMAR)
Mean CPS (logMAR)
Colenbrander logRAD/logMAR ratio (%)
Difference in logRAD/logMAR ratio (%)
Mean logRAD/logMAR ratio (%)

b)


Radner reading acuity (logRAD)
Difference in reading acuity (logRAD)
Mean reading acuity (logRAD)
Radner MRS (wpm)
Difference in MRS (wpm)
Mean MRS (wpm)
Radner CPS (logMAR)
Difference in CPS (logMAR)
Mean CPS (logMAR)
Radner logRAD/logMAR ratio (%)
Difference in logRAD/logMAR ratio (%)
Mean logRAD/logMAR ratio (%)

c)

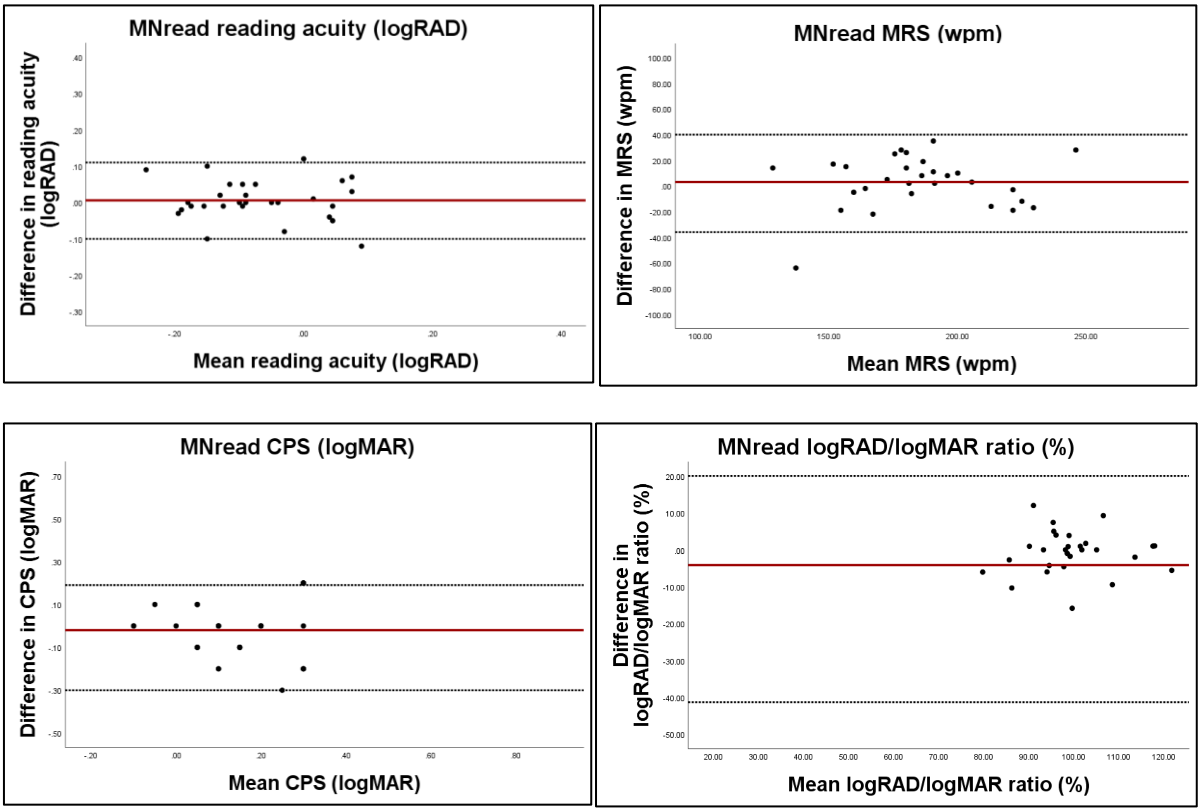


Figure 4.2: Test-retest reliability: Bland-Altman plot of the differences between the two testing sessions (First minus second) in presbyopic subjects. In (a) Bland-Altman plots for reading performance metrics in the Colenbrander chart showed high agreement between the two testing sessions for MRS, CPS and LogRAD/LogMAR ratio, but not for RA; (b) Bland-Altman plots for reading performance metrics in the Radner chart showed high agreements in all reading performance metrics; (c) Bland-Altman plots for the MNread chart showed a low agreement in logRAD/logMAR ratio, but high agreement for RA, MRS and CPS (n = 29).

The results of the Bland-Altman agreement analysis are shown in Figure 4.2. In the Radner reading chart, although the differences between the two testing sessions for all the reading performance metrics were statistically significant, high agreements were found. In the Colenbrander chart, low agreement was found for measuring the reading acuity while the agreements were high for the other metrics. Also, in the MNread chart, high Bland-Altman agreements were found for all the reading performance metrics except for the logRAD/logMAR ratio where the agreement was low.

4.2.1.2 Repeatability of Bailey-Lovie unrelated words reading chart

The Pearson correlation coefficient showed that RA (r = 0.574), CPS (r = 0.550) and logRAD/logMAR ratio were moderately correlated between the two testing sessions. The correlation for measuring the MRS was higher (r = 0.713). No statistically significant differences were found for measuring all the reading performance metrics between both sessions and the agreements were high (Fig. 4.3).

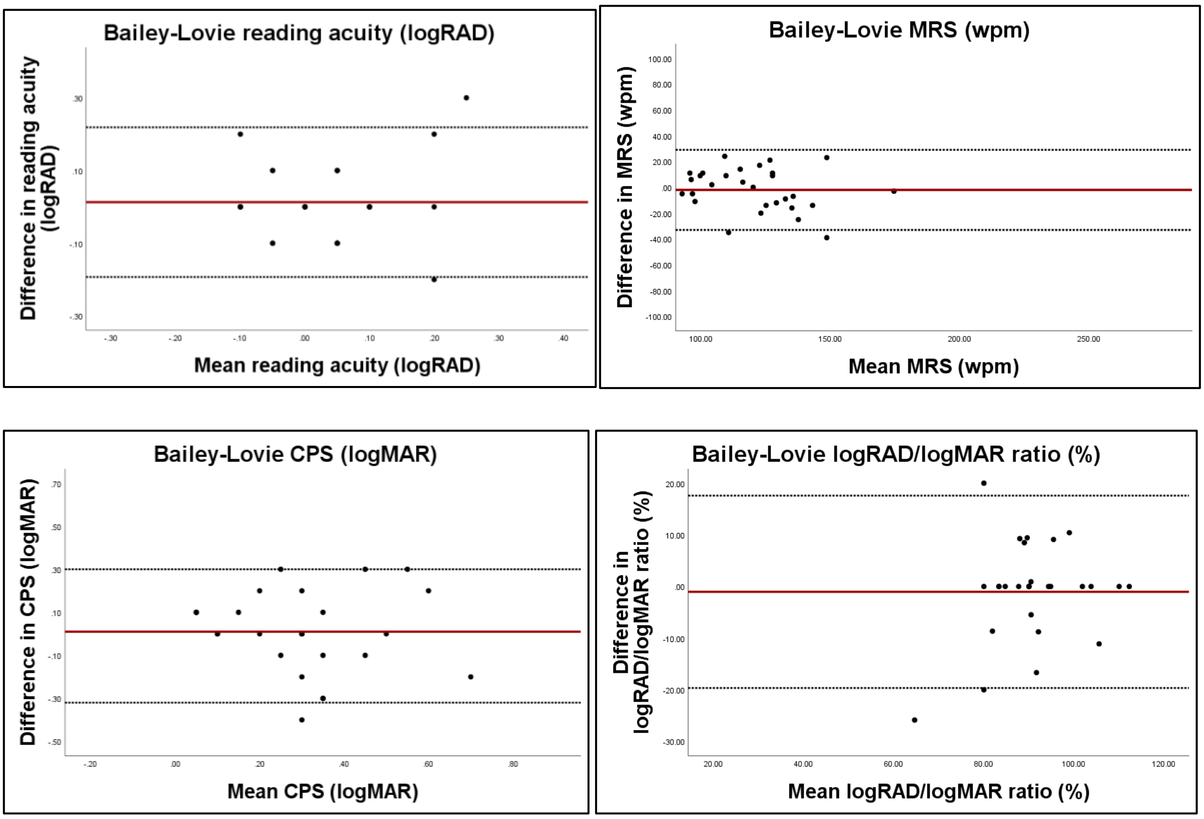


Figure 4.3: Test-retest reliability Bailey-Lovie words chart in presbyopic subjects: Bland-Altman plots of the differences between the two testing sessions (first minus second) for RA, MRS, CPS and LogRAD/LogMAR ratio. High agreements were found for all reading performance metrics.

4.2.1.3 Repeatability of Times New Roman chart (Ciba Vision reading card)

The older age participants’ ability to read the smallest print size in the reading charts was varied. However, in TNR charts all the participants were able to read the smallest print size (N6) in the two testing sessions. There was a strong correlation (*r* = 0.818) and agreement between the first and second session for measuring MRS without significant difference. The Bland-Altman agreement was low for measuring CPS.

4.2.1.4 Repeatability of International Reading Speed test (IReST)

Since the only metrics that can be calculated from IReST are the average reading speed. The average reading speed measurement for the three paragraphs selected (text number six, seven and eight) was strongly correlated between the two testing sessions (r = 0.941). Bland-Altman evaluated a good agreement in the repeatability of average reading speed measurements (Fig. 4.4). Table 4.3 summarises the repeatability results of the reading performance metrics for the six reading charts.

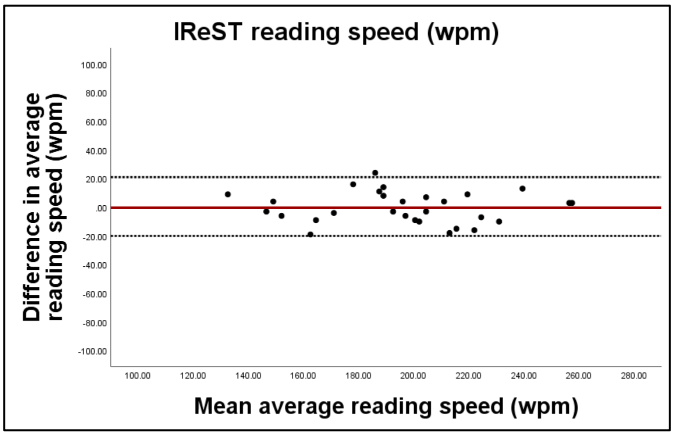


Figure 4.4: Test-retest reliability of the IReST chart in presbyopic subjects: Bland-Altman plots of the differences between the two testing sessions (first minus second) for average reading speed showing a high agreement.

The coefficients of repeatability of the reading performance metrics in the proposed reading test charts are presented in Table 4.2. The Colenbrander reading chart yielded the highest CR (less repeatable) for MRS and CPS. For the logRAD/logMAR ratio, the MNread chart yielded the highest CR (9.70). The IReST reading chart showed the lowest CR (more repeatable) for MRS compared to the other reading charts in this group of elderly participants.

Table 4.2: The coefficient of repeatability for the reading performance metrics between two testing sessions using different reading charts in prespyopic subjects

Reading metrics	(CR) Colenbrander	(CR) Radner	(CR) MNread	(CR) Bailey-Lovie	(CR) TNR	(CR) IReST
RA (logRAD)	0.10	0.11	0.11	0.21	-	-
MRS (wpm)	56	36	40	32	37	20
CPS (logMAR)	0.30	0.22	0.26	0.34	-	-
logRAD/logMAR ratio (%)	10.6	11.2	37.8	19.08	-	-

Table 4.3: Summary of the repeatability results of the reading performance metrics for the six reading charts in presbyopic subjects

Reading chart	Reading performance metrics	Correlation coefficient (r value)	Differences in measurements (t test) *p* value	Bland-Altman limits of agreement
Colenbrander	RA (logRAD)	0.620	0.096	[0.11; -0.07]
	MRS (wpm)	0.560	0.673	[52; -56]
	CPS (logMAR)	0.599	0.477	[0.31; -0.27]
	LogRAD/LogMAR ratio (%)	0.746	0.089	[8.8; -12.0]
Radner	RA (logRAD)	0.805	0.007	[0.13; -0.06]
	MRS (wpm)	0.781	0.028	[45; -20]
	CPS (logMAR)	0.638	0.023	[0.26; -0.16]
	LogRAD/LogMAR ratio (%)	0.807	0.009	[8.14; -14.1]
MNread	RA (logRAD)	0.845	0.551	[0.11; -0.10]
	MRS (wpm)	0.767	0.441	[40; -36]
	CPS (logMAR)	0.551	0.284	[0.19; -0.30]
	LogRAD/LogMAR ratio (%)	0.420	0.252	[20.0; -41.3]
Bailey-Lovie	RA (logRAD)	0.574	0.489	[0.22; -0.19]
	MRS (wpm)	0.713	0.621	[29; -33]
	CPS (logMAR)	0.550	0.757	[0.30; -0.32]
	LogRAD/LogMAR ratio (%)	0.645	0.571	[17.5; -19.6]

Times New Roman	MRS (wpm)	0.818	0.969	[37; -35]
	CPS (N notation)	0.327	0.663	[0.85; -0.79]
IReST	Average reading speed (wpm)	0.941	0.880	[21; -20]

4.2.2 Reliability

Reliability or inter-chart reliability of the reading charts was determined by calculating the Cronbach's alpha coefficient of the reading performance metrics measurements obtained from each reading chart at the two testing sessions in the elderly subjects. In the Colenbrander reading chart, the Cronbach's alpha resulted in higher reliability result for measuring loRAD/logMAR ratio (α = 0.845) followed by MRS (α = 0.783), CPS (α = 0.748) and then RA (α = 0.727). Good reliability results of all the reading performance metrics were found in the Radner reading chart, ranging from 0.779 to 0.893. In the MNread chart, very high reliability was found for measuring RA (α = 0.915). However, poor reliability results were found for measuring logRAD/logMAR ratio (α = 0.492). Good Cronbach's alpha coefficients resulted for measuring the reading performance metrics in the Bailey-Lovie word reading chart. However, the reliability was higher for measuring MRS (α = 0.825) compared to other reading metrics in the Bailey-Lovie chart. Reading acuity measured by the TNR chart was identical for all the presbyopic subjects. Therefore, the reading acuity measurements' reliability cannot be computed because the variable was constant. Times New Roman chart resulted in an excellent reliability for measuring MRS (α = 0.90) and very poor reliability for measuring CPS (α = 0.323). In the IReST reading chart, very high Cronbach's alpha was found for measuring the average reading speed in presbyopic subjects (α = 0.970). Table 4.4 summarises the reliability results of the reading charts.

Table 4.4: The Cronbach's alpha coefficient results for evaluating the reliability of the reading chart in presbyopic subjects.

Reading chart	Reading performance metrics	Cronbach's alpha	Reading chart	Reading performance metrics	Cronbach's alpha
Colenbrander	RA (logRAD)	0.727	**Bailey-Lovie**	RA (logRAD)	0.727
	MRS (wpm)	0.783		MRS (wpm)	0.825
	CPS (logMAR)	0.748		CPS (logMAR)	0.709
	LogRAD/LogMAR ratio (%)	0.854		LogRAD/LogMAR ratio (%)	0.776
Radner	RA (logRAD)	0.891	**TNR**	MRS (wpm)	0.90
	MRS (wpm)	0.870		CPS (N notation)	0.323
	CPS (logMAR)	0.779			
	LogRAD/LogMAR ratio (%)	0.893	**IReST**	Average reading speed (wpm)	0.970
MNread	RA (logRAD)	0.915			
	MRS (wpm)	0.867			
	CPS (logMAR)	0.710			
	LogRAD/LogMAR ratio (%)	0.492			

4.2.3 Comparison of the reading errors between the charts

The averages of the reading errors made in each chart in each session were calculated. No statistically significant differences were found in the number of reading errors in all reading test charts between the first and the second testing sessions (*P* <0.05). The single significant difference between the two testing sessions was found in the TNR chart (*P* = 0.028), where the presbyopic participants made fewer reading errors in the second test session compared to the first one. This difference was 1.17 ± 2.73, 95% confidence interval CI: (0.135; 2.21). Comparing the reading errors between the reading charts in the first visit showed that the Bailey-Lovie word reading chart resulted in fewer reading errors compared to the all other reading charts. The differences were as follows: between Bailey-Lovie and Colenbrander (2.44 ± 3.01) *P* = 0.000, Bailey-Lovie and Radner (2.55 ± 2.78) *P* = 0.000, Bailey-Lovie and MNread (2.0 ± 3.04) *P* = 0.001, Bailey-Lovie and TNR (2.03 ± 3.04) P = 0.004 and between Bailey-Lovie and IReST (1.86 ± 4.61) *P* = 0.038. No significant differences were found in a number of reading errors between all the other reading charts.

Table 4.5: Number of reading errors per reading chart

Test session	**Colenbrander**	**Radner**	**MNread**	**Bailey-Lovie**	**Times New Roman**	**IReST**
First (mean± SD)	4.10 ± 2.73	4.21± 2.70	3.65± 3.25	1.65± 1.70	3.69± 3.85	3.51± 5.09
Second (mean± SD)	4.17 ± 4.31	4.24± 2.40	3.37± 2.62	2.03± 2.23	2.51± 2.96	3.17± 3.16

4.2.4 Comparison of reading performance outcomes between reading test charts in presbyopic subjects

The data were normally distributed and Pearson correlation coefficient and Bland-Altman agreement together with a Student t test were performed to evaluate and compare the reading test charts (Colenbrander, MNread, Radner and Bailey-Lovie word reading chart) for measuring the reading performance metrics in presbyopic participants during the first testing session. *P* <0.05 was considered statistically significant. The IReST reading test and TNR chart were excluded from this comparison since the IReST uses only one print size. The exclusion of the TNR chart was based on the results of Chapter Three which showed that making the comparison between the paragraphs, as in the TNR, and the sentences as in the other chart is not reasonable.

From the results in Table 4.6 below, overall it is apparent that different reading performance outcomes were obtained when using different reading test charts in presbyopic subjects. For measuring RA, statistically significant differences with low Bland-Altman agreements were found between (Colenbrander and Radner), (Colenbrander and Bailey-Lovie) and (MNread and Bailey-Lovie). These differences were more than one line of reading acuity in logMAR. Also, statistically significant difference was found between Radner and MNread; this difference was in RA LogRAD: 0.08 ± 0.06, 95% confidence interval CI: (0.05; 0.10) $P = 0.000$. However, the correlation ($r = 0.789$) and the agreement were high between Radner and MNread. Moderate correlation and low agreement were found between Colenbrander and MNread, but t-test found no statistically significant difference. The results showed that no statistically significant differences for measuring RA were found between Radner and Bailey-Lovie with good correlation and agreement (Fig. 4.5).

The correlation between MRS measurements between the reading charts ranged from weak ($r = 0.280$) to moderate correlation ($r = 0.596$) with statistically significant differences, $P < 0.05$. However, the correlation for measuring MRS was relatively high between Radner and MNread ($r = 0.738$) with no significant difference. The Bailey-Lovie word reading chart always resulted in slower reading speed compared to Colenbrander, Radner and MNread, while the Colenbrander chart resulted in faster reading speed compared to the other charts in presbyopic subjects (Fig. 4.6). The Bland-Altman agreements were low between Bailey-Lovie chart and the other charts. On the other hand, the agreements were good between the continuous text reading charts (Radner, Colenbrander and MNread) for measuring the MRS, see Figure 4.7.

Table 4.6: The correlation and differences together with Bland-Altman limit of agreement between the reading charts within the same test session in presbyopic subjects. The differences in the reading performance metrics between the charts showed a variation of the results.

Reading charts	***Reading performance metrics***	***Correlation coefficient (r value)***	***Differences in measurements t- test p value (mean ± SD)***	***Bland-Altman agreement***
Colenbrander + Radner	RA (logRAD)	0.627	*P* = 0.000 (-0.06± 0.07)	[0.20; -0.07]
	MRS (wpm)	0.560	*P* = 0.000 (47± 28)	[101; -8]
	CPS (logMAR)	0.212	*P* = 0.095 (-0.06± 0.20)	[0.33; -0.40]
	LogRAD/LogMAR ratio (%)	0.678	*P* = 0.000 (6.51± 6.78)	[19.7; -5.0]
Colenbrander + MNread	RA (logRAD)	0.620	*P* = 0.414 (-0.011± 0.076)	[0.15; -0.14]
	MRS (wpm)	0.509	*P* = 0.000 (40± 30)	[98; -18]
	CPS (logMAR)	0.401	*P* = 0.045 (0.06±0.16)	[0.37; -0.2]
	LogRAD/LogMAR ratio (%)	0.320	*P* = 0.497 (2.52± 19.72)	[41.1; -20.0]

Table 4.6: (continued)

Reading charts	***Reading performance metrics***	***Correlation coefficient (r value)***	***Differences in measurements t- test p value (mean ± SD)***	***Bland-Altman agreement***
Colenbrander + Bailey-Lovie	RA (logRAD)	0.498	*P* = 0.000 (-0.08± 0.10)	[0.1; -0.27]
	MRS (wpm)	0.280	*P* = 0.000 (106± 31)	[166; 45]
	CPS (logMAR)	0.251	*P* = 0.011 (-0.10± 0.21)	[0.31; -0.51]
	LogRAD/LogMAR ratio (%)	0.630	*P* = 0.000 (8.23± 9.33)	[26.5; -10.0]
Radner+ MNread	RA (logRAD)	0.789	*P* = 0.00 (0.08± 0.06)	[0.19; -0.03]
	MRS (wpm)	0.738	*P* = 0.101 (7± 22)	[40; -36]
	CPS (logMAR)	0.697	*P* = 0.000 (0.012±0.10)	[0.30; -0.07]
	LogRAD/LogMAR ratio (%)	0.522	*P* = 0.240 (3.99± 17.88)	[31.0; -20.0]

Table 4.6: (continued)

Reading charts	***Reading performance metrics***	***Correlation coefficient (r value)***	***Differences in measurements t- test p value (mean ± SD)***	***Bland-Altman agreement***
Radner + Bailey-Lovie	RA (logRAD)	0.661	*P* = 0.252 (-0.10± 0.09)	[0.18; -0.16]
	MRS (wpm)	0.661	*P* = 0.000 (59± 22)	[102; 20]
	CPS (logMAR)	0.415	*P* = 0.217 (-0.04± 0.17)	[0.29; -0.37]
	LogRAD/LogMAR ratio (%)	0.722	*P* = 0.276 (1.17± 8.32)	[17.4; -15.2]
Bailey-Lovie + MNread	RA (logRAD)	0.744	*P* = 0.000 (-0.10± 0.08)	[0.05; -0.25]
	MRS (wpm)	0.596	*P* = 0.000 (66± 24)	[113; 18]
	CPS (logMAR)	0.571	*P* = 0.000 (-0.16±0.15)	[0.10; -0.45]
	LogRAD/LogMAR ratio (%)	0.247	*P* = 0.161 (5.70± 12.35)	[25.0; -36.0]

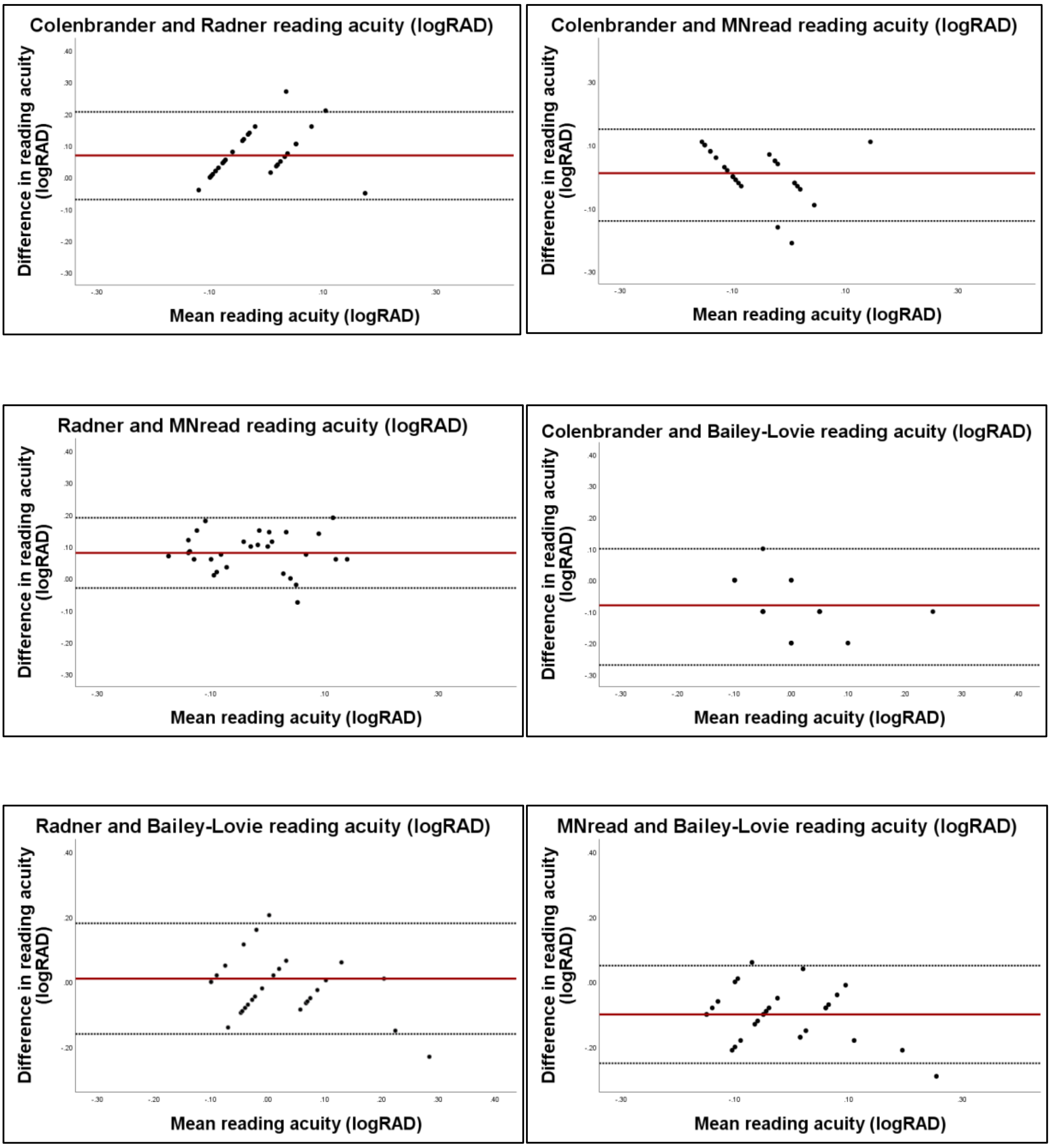


Figure 4.5: Bland-Altman agreement plots of the reading acuity between the charts in presbyopic subjects. The differences of the mean between the charts is plotted against the mean reading acuity of both charts

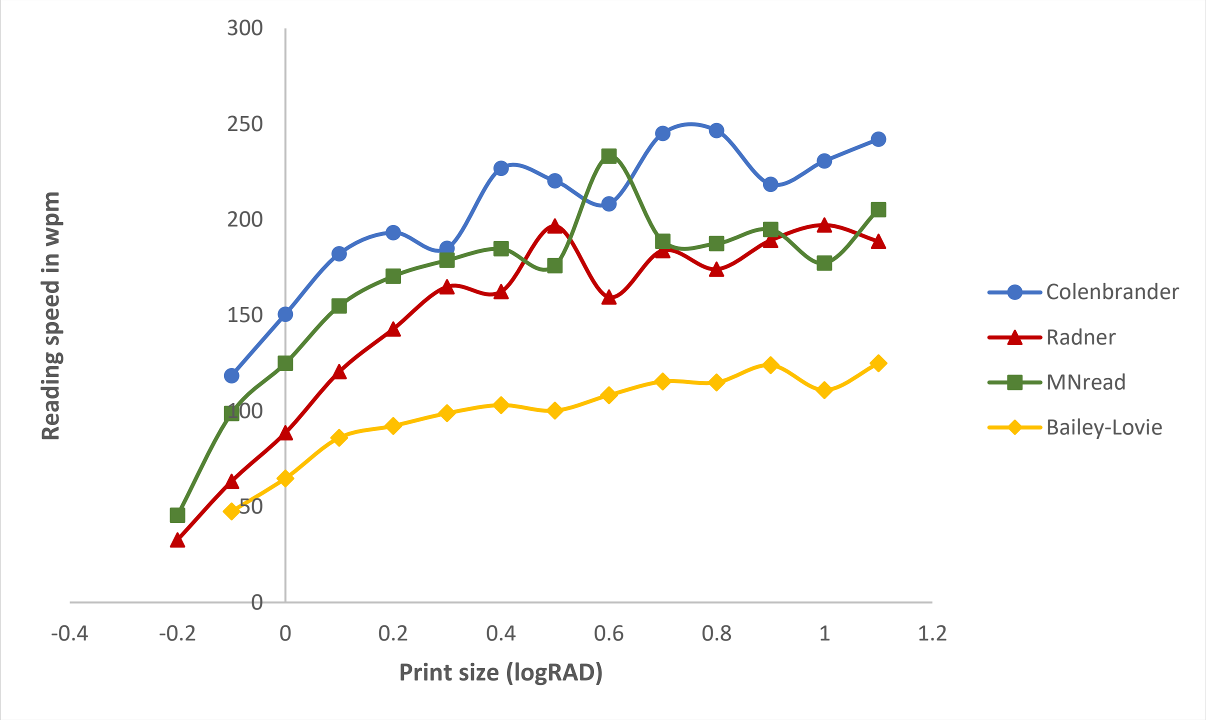


Figure 4.6: Mean reading speed in wpm for 29 presbyopic participants plotted as a function of print size for Colenbrander, Radner, MNread and Bailey-Lovie

A comparison of the CPS between the reading charts revealed poor Pearson correlations coefficients, ranging from r = 0.212 to r = 0.571 except between the Radner and MNread where the correlation was stronger (r = 0.697). However, the agreement between Radner and MNread was low, with statistically significant differences. Although the Bland-Altman agreements were high for measuring CPS between Colenbrander and MNread and Colenbrander and Bailey-Lovie, statistically significant differences were found *P* <0.05. The correlation and agreement were very weak between MNread and Bailey for measuring CPS (Fig. 4.8).

As shown in Table 4.6, the Bland-Altman agreements for measuring the logRAD/logMAR ratio in same testing session were low and varied between all the charts except between Colenbarander and Rdaner where the agreement was high (Fig 4.9). Although the agreement was high between Colenbrander and Radner, t-test found a statistically sginificant difference. This difference was in LogRAD/LogMAR ratio (%): 6.51± 6.78, 95% confidence interval CI: (3.39; 9.06).

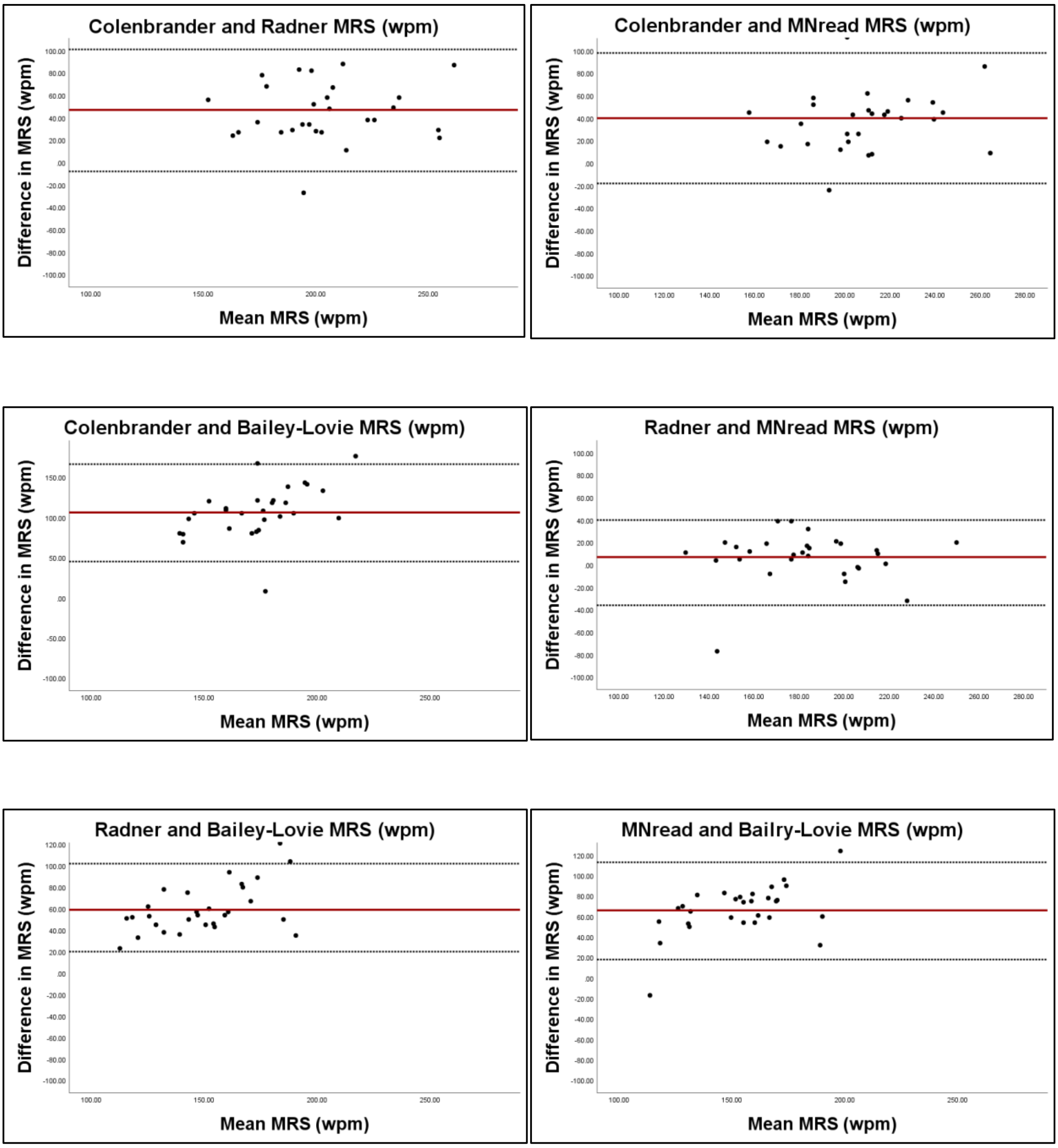


Figure 4.7: Bland-Altman agreement plots of the reading speed between the charts in presbyopic subjects. The differences of the mean between the charts is plotted against the mean reading speed of both charts

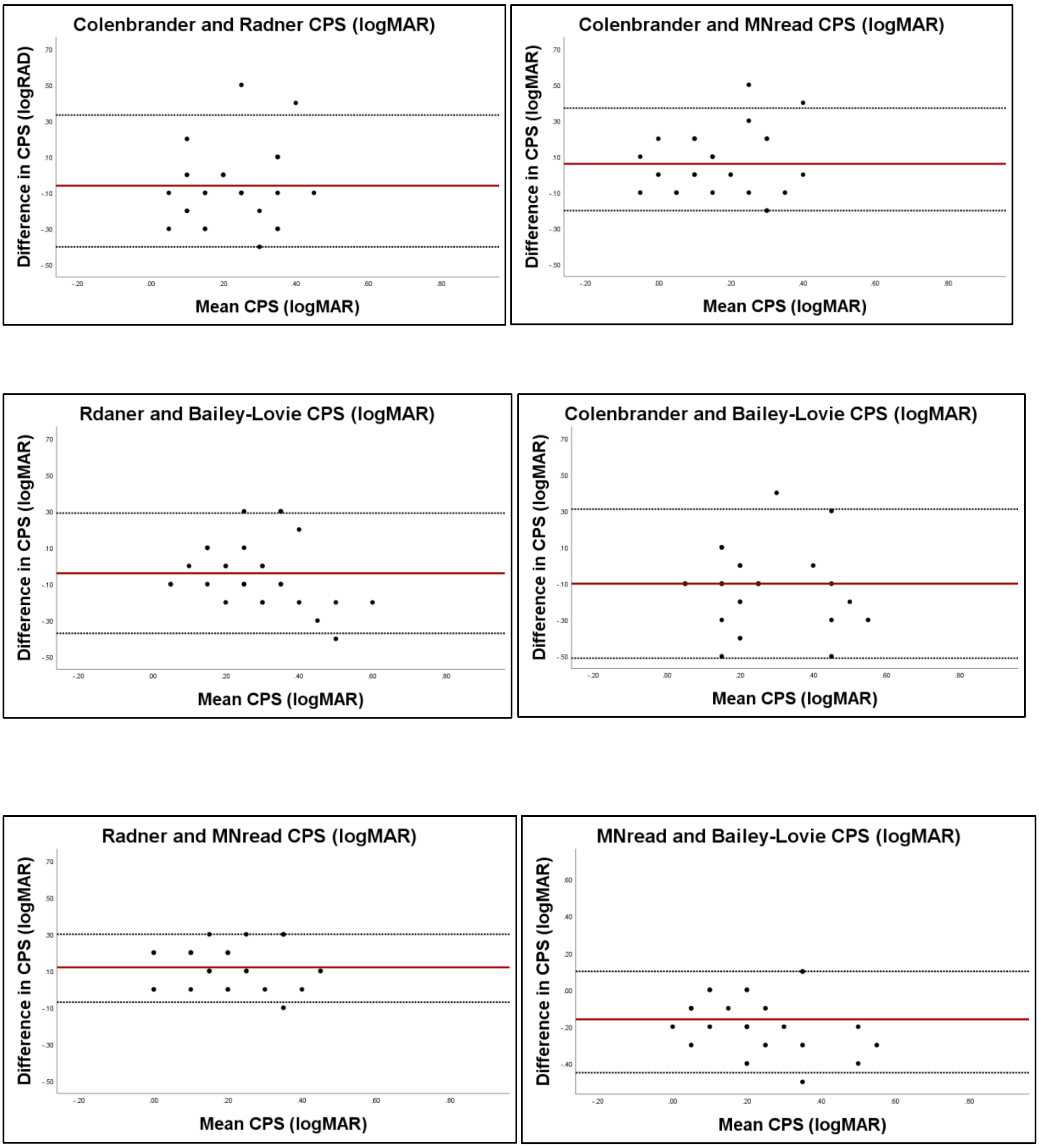


Figure 4.8: Bland-Altman agreement plots of the CPS between the charts in presbyopic subjects. The differences of the mean between the charts are plotted against the mean CPS of both charts.

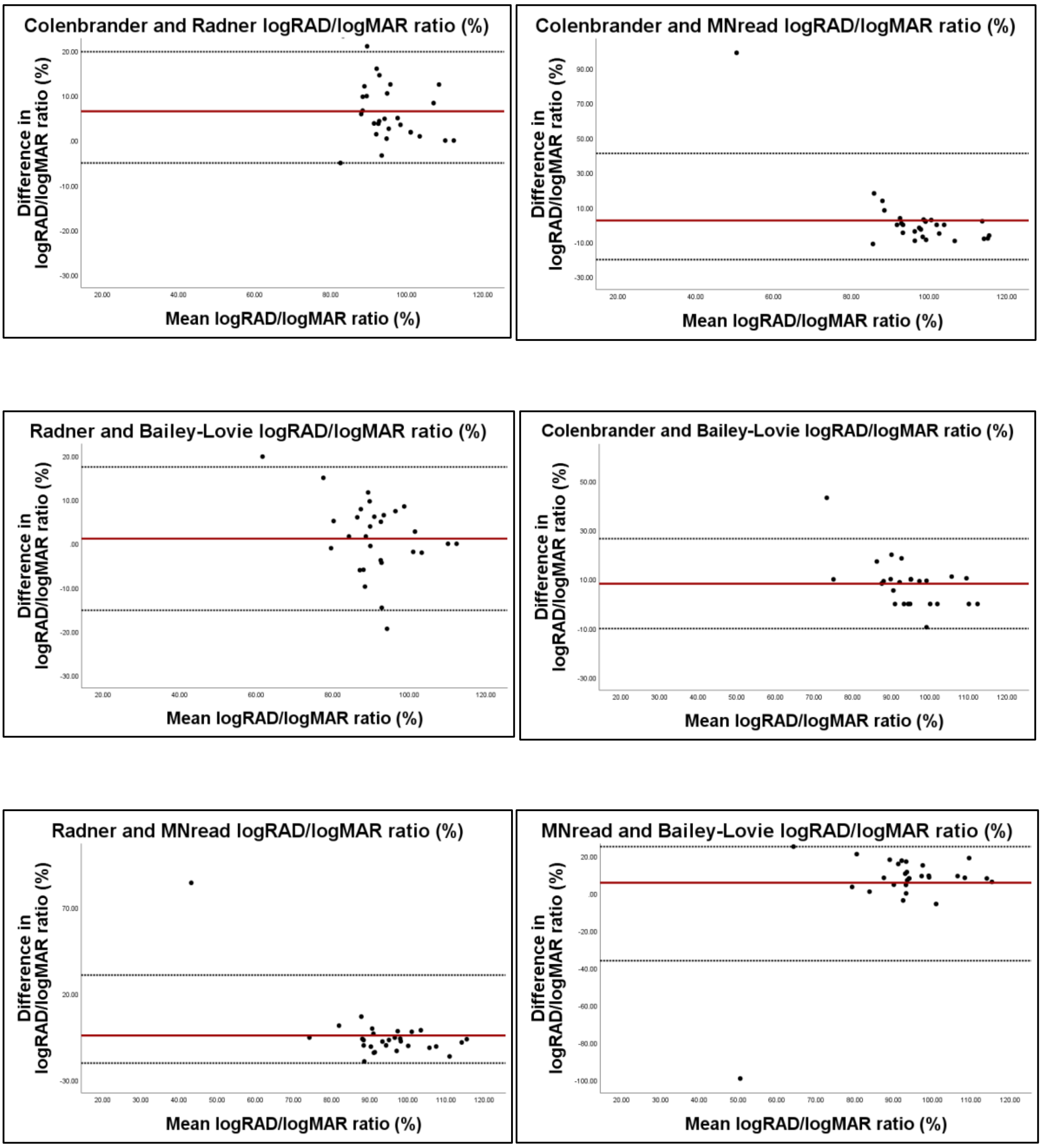


Figure 4.9: Bland-Altman agreement plots of the logRAD/logMAR ratio between the charts in presbyopic subjects. The differences of the mean between the charts are plotted against the mean logRAD/logMAR ratio of both charts.

4.3 Discussion

4.3.1 Repeatability and inter-chart reliability of the reading test charts for measuring the reading performance metrics in people with presbyopia

Evaluation of the reading test charts has a clinical importance for people with presbyopia as a pre- or post-operative measure and for optical correction using spectacles or multifocal contact lenses. For research and clinical examination, reading performance analysis should provide reliable and repeatable results.

Because, in reviewing the literature, no data were found on direct association and differences between the available reading test charts in the English language on people with presbyopia tested under the same clinical condition, the goal of this chapter was to analyse and compare the repeatability and inter-chart reliability of the reading performance metrics measured with different reading test charts in older participants.

In this study, one of the criteria for selecting the presbyopic subjects was having at least A-level education. (Munch et al., 2016) confirmed that the educational background of the patients had an influence on the reading performance measurements. They reported that the reading speed measurements for medical student participants were significantly faster than that for blue-collar factory workers. Therefore, avoiding varying educational background of the participants should be considered when making the comparison between the reading test charts.

The findings showed that the MNread test chart had good repeatability and reliability for measuring the reading performance metrics (RA, MRS and CPS). The reading acuity measurements had the highest repeatability (r = 0.845) and reliability (α = 0.915). These findings are consistent with that of (Virgili et al., 2004a). Another study by (Mataftsi et al., 2013) reported high repeatability results for the MNread chart with adults age range 25 to 59 years and children aged eight years. However, their results were not applicable to the real World because the time interval between the two sessions was only 30 minutes. Evaluation of the repeated measurements taken in this short period of time could increase the learning and practice effect and lead to higher repeatability results. In addition, taking the reading performance measurements twice on the same day may not apply to clinical practice setting as the patients are followed up longitudinally. In the present study, time interval of two to three weeks was used to evaluate the repeatability of the reading charts. However, poor correlation (r = 0.420) and Bland-Altman agreement between the two testing sessions were found for measuring logRAD/logMAR ratio and this result was confirmed by the higher CR (37.8). Also, the Cronbach's alpha yielded poor reliability (α = 0.492) in the MNread chart for measuring logRAD/logMAR ratio in presbyopic subjects.

The results of the MNread test chart estimated 95% coefficient of repeatability of 0.11 logRAD for RA, 0.26 logMAR for CPS and 40 wpm for MRS at the two visits. These values are comparable with those reported in (Subramanian and Pardhan, 2009, Patel et al., 2011) studies except for MRS the CR was much higher. The reason of this discrepancy is attributed to the patient-related factors since Subrmanian and Pardhan (2009) investigated the repeatability of the MNread chart in subjects with impaired vision, while Patel et al. (2011) investigated age-related macular degeneration. The possible interference of the disease progression could increase the variability reported in the repeated measurements.

Prior studies that have evaluated the IReST reading charts were performed in young, healthy population. The IReST reading chart was developed for measuring the reading speed not CPS or reading acuity using 1M print size. The results showed that the IReST reading chart has very good characteristics in terms of repeatability (r = 0.941) and inter-chart reliability (α = 0.970) with older age population. The IReST paragraphs contain multiple sentences equal in linguistics, which influences the predictability of the text and result in less noise and variability for measuring the reading speed. In addition to that, the reading time for paragraphs is longer than for a sentence and this could contribute to lower the variance between the reading speed measurements (Altpeter et al., 2015). Reading the longer text paragraphs represents the real-life reading performance as the reading task occurs over a long period of time (Stangler-Zuschrott, 1990).

In this presbyopic group of subjects, the test-retest repeatability for the Colenbrander reading chart resulted in a low Bland-Altman agreement for measuring reading acuity and lower Pearson correlation for measuring MRS compared to the other reading charts. This also accords with our earlier observations with a younger age group of subjects. As was pointed out in the methods section in the previous chapter, the method of calculating the reading errors in the Colenbrander chart in terms of the word value is not as clear as in the Radner (0.005 logRAD per incorrect read syllable) and MNread (0.1 logRAD per incorrect read word) charts. This result of the poor agreement for measuring RA in the Colenbrander chart may be explained by the fact that the scoring "weight" of the incorrect read word is lower compared to MNread and Radner and this may significantly influence the total reading acuity score.

In the Radner reading chart, the outcomes of the CR are higher compared to Maaijwee et al. (2008b) study; however, they tested older age group with age 50 years or older affected by a macular disease. Their coefficients of repeatability were 8.2 wpm for reading speed, 0.036 logRAD for reading acuity and 0.06 logMAR for CPS compared to this study's results of 36 wpm, 0.11 logRAD and 0.22 logMAR for CPS. However, one criticism of their study is that they measured the reading speed two consecutive times and then extra measurement was made if there was a difference more than 0.2s, which was found to decrease the variability of the results. Although the Bland-Altman agreements were high in the Radner chart, the t-test showed a statistically significant difference (P <0.05) between the two sessions for measuring all the reading performance metrics.

Turning now to the Bailey-Lovie word reading chart, as indicated earlier, this chart has never been investigated in terms of test-retest and inter-chart reliability. The findings have demonstrated that the Bailey-Lovie reading chart could provide a highly repeatable and reliable measure of the reading performance. Bland-Altman agreements were high for measuring all the reading performance metrics without any statistically significant differences. In addition, the Cronbach's alpha ranged from 0.709 to 0.825. Bailey and Lovie (1980) designed the word reading chart taking care to avoid syntactic association between the words so that the reading performance can be measured with greater reliability. Based on the principle of the Bailey-Lovie reading chart, some modern reading charts have been developed to exclude the linguistic aspect, such as the SKread vision test (MacKeben et al., 2015). The SKread vision test combines the format of the MNread chart with a sequence of random letters and words.

4.3.2 Comparison of the reading errors between the charts

Although in most reading test charts around four reading errors were counted, the number of errors in the Bailey-Lovie word chart was significantly fewer. The features of the Bailey-Lovie unrelated word chart must be responsible for that. This also accords in the young participants as explained earlier in Chapter Three. The young and the older participants in both studies were normally sighted. However, this outcome is contrary to that of (Nair et al., 2006) who found that people with eye disease made more reading errors with an unrelated word chart compared to the continuous text reading charts. To develop a full picture, this comparison of the reading errors between random word and continuous text charts will be investigated in the next chapter for people with cataracts to address whether the status of the eye disease could change the results.

Another significant aspect, in the TNR chart only, is that the presbyopic participants made fewer reading errors in the second testing session compared to the first session. It seems that the learning and the practice effect plays an important role on the result of the TNR chart. This result is likely to be related to the complexity and the reading level of the text. The TNR chart uses some difficult words like “astigmatism”, “varifocal glasses” and “bifocals”, and this seems to affect the level of the test sentences in terms of difficulty and the amount of the time needed to read the sentence. Although, the participants’ level of education was good, using words that are not familiar to the public could affect the result of the reading performance.

4.3.3 Comparison of reading performance outcomes between reading test charts in presbyopic subjects

The outcomes of reading performance metrics between the reading test at the same testing session charts were varied in older age participants.

The Radner and MNread charts have similar design in the way that they provide for measuring the reading performance metrics. However, the direct comparison between them in this old age group of subjects revealed statistically significant differences for measuring reading acuity, CPS and low agreement for measuring the logRAD/logMAR/ratio. According to the comparison of MRS between Radner and MNread, what is striking in Figure 4.6 is the high discrepancy of MRS measurement in 0.6 logMAR letter size. The 0.6 logMAR letter size sentence in the MNread chart’s “My mother loves to hear the young girl sing in the morning” noticeably read faster than 0.6 logMAR letter size in the Radner chart’s “Old Stephen wanted to repair his snowmobile which was older than everyone thought”. This finding was also noticed with young subjects in the previous chapter. The discrepancy could be attributed to the number of syllables of the word used in both charts. In this exact letter size, the Radner chart used two words with three syllables and four words with two syllables, while the MNread chart used two words with two syllables (mother and morning) and the other words have only one syllable. The use of the short words makes the MNread chart score faster reading speed compared to the Radner.

As shown in Figure 4.6 and Table 4.6, the Bailey-Lovie word reading chart resulted in significantly slower MRS compared to the Radner, MNread and Colenbrander with low Bland-Altman agreements. On the other hand, the Colenbrander reading chart revealed faster reading speed compared to the other charts. The use of unrelated syntactic between the words made the reading more difficult and slowed down compared to the continuous test reading chart (SASS et al., 2006). SASS et al. (2006) confirmed that the removal of the sentence context had a greater impact on reducing the reading speed irrespective of the age. This finding is similar to that reported for the younger age group in the previous chapter. Also, the approach used in this investigation is similar to that found by other researchers. Legge et al. (2002) exhibited a 25% increase in reading speed related to context while Morton (1964) found 33% increase in reading speed. Another study by Bullimore and Bailey (1995) reported similar context advantages for increasing the reading speed for normally sighted people. The prediction of the next word is an important factor in enhancing faster reading speed, as in the continuous text reading charts.

In the assessment of the reading abilities, reading acuity is the most important indicator of the visual capacity of reading. One question in the previous study sought to determine if the variability of the reading acuity measurements are influenced by age. In this group of subjects, again the reading acuity measurements are not well correlated between the reading charts (Colenbrander, Radner, MNread and Bailey-Lovie) with differences of more than one line in logMAR. It has been demonstrated earlier that the smallest print sizes are varied between the charts. The smallest print size in the Colenbrander reading chart is (-0.1 logMAR) while 12 presbyopic participants were able to read (-0.2 logMAR) letter size. These differences could be a possible source of the variation on reading acuity outcomes. In this study, the same scoring rule was used for all the reading charts regarding the end point for reading acuity as reading more than 80% of the words in the sentences accurately (Stifter et al., 2005c).

In summary, this study was designed to determine the repeatability and reliability of reading performance metrics using different reading charts in addition to comparing the reading performance outcomes in people with presbyopia. The values obtained from the result of this study may be used to give insight into future clinical trial design and may be used to guide optometrists to use uniform and standardized reading test charts in clinical practice. Since some ocular diseases, such as cataract, decrease reading ability, the feasibility of reading charts for patients with cataract is of interest and will be investigated in the next chapter.

4.4 Conclusion

To conclude this chapter, the simultaneous determination of the RA, MRS, CPS and LogRAD/LogMAR ratio in the same examination using a standardised reading test chart is important in the diagnosis of reading performance in people with presbyopia. One of the principle findings of this research was that the difference in reading performance measurements depends on the used reading test material. In the present study of an older group, taking into account the variation between and within the subjects and the study setting, it considered the strength of the reading charts that provide reliable reading performance results. The Bailey-Lovie word reading chart showed higher inter-chart and test-retest reliability for measuring the reading performance metrics. Although, the MNread chart has low reliability and agreement for measuring logRAD/logMAR ratio, it has a strong reliability and repeatability for measuring the other reading performance metrics, especially the reading acuity.

Chapter 5

Reliability and repeatability of the reading performance metrics in different reading charts for subjects with cataract

5 Introduction

As a result of the rapid increase in the mean age of the population, the number of people with presbyopia will grow, along with the number of patients with age-related ocular diseases, such as cataract. The use of reading test in patients with cataract will become even more important. This is problematic as there is no consensus regarding which reading test provides the best results in terms of test-retest and inter-chart reliability.

Previous studies suggest that using an optimal reading test has a value in patients with cataract to assess vision before and after surgery (Fine and Rubin, 1999, Elliott et al., 2001). It has been proven that there is a strong influence of the morphological type of cataract on reading performance. Patients with posterior subcapsular cataract had a significant reduction in MRS while those with nuclear cataract achieved normal MRS preoperatively (Stifter et al., 2004b). Therefore, including a standardized and calibrated reading test in clinical examination may facilitate the management. Also, it has been found that the use of standardized reading test has the ability to discriminate the visual impairment caused by cataract, maculopathy or ARMD while distance visual acuities were comparable between these ocular diseases and underestimated the patient’s functional impairment (Stifter et al., 2005c).

The present study evaluated the test-retest and inter-chart reliability of the reading test charts, Colenbrander, Radner, MNread, Bailey-Lovie, IReST and TNR, for measuring the reading performance metrics, including RA, MRS, CPS and logRAD/logMAR ratio for patients with cataract.

5.1 Methods

Cataract patients were recruited from the Aston University Eye Clinic. Ethical approval was obtained from Aston University Ethical Committee (see Appendix 3) in accordance with the Helsinki Declaration and all the subjects gave their consent to participate in the study.

The specific exclusion criteria were as follows:

- Previous ocular surgery
- Presence of diabetes mellitus
- Presence of abnormality or ocular disease other than cataract
- Presence of binocular vision problems
- Use of any medications that could affect the reading ability

The specific inclusion criteria were as follows:

- Presence of cataract, as defined below
- Native English speakers
- Level of education at or above A-level (secondary school year 13)

The Lens Opacities Classification System (LOCS) III was used to categorise and grade cataract using the slit-lamp (Chylack et al., 1993). This grading system has been widely used in clinical research (Davison and Chylack, 2003). The presence of cataract was defined as more than grade two nuclear colour or opalescence, more than grade one cortical opacity, and any sign of subcapsular opacity (Elliott et al., 2001).

Reading performance measurements were performed monocularly in the eye with cataract with optimal refractive correction that compensated for a testing distance of 40cm. The test-retest and inter-chart reliability of the proposed reading charts (Colenbrander, Radner, MNread, Bailey-Lovie, IReST and TNR) were evaluated in cataract patients at two testing sessions with a time interval of two to three weeks. The same refractive correction was used in the two testing sessions. Although, the time interval was short for cataract progression, it was important to ensure that the status of cataract was stable by performing a slit-lamp examination at the second visit.

The material, measurement procedure, statistical analysis and the methods of calculating the reading performance metrics were the same as in Chapter Three and Four. In this chapter, further analysis, using multiple regression analysis expressed as coefficient of determination r^2 was conducted to determine to what degree the variability of the type and grade of cataract accounted for the reading performance metrics for each reading chart.

5.1.1 Results

Twenty-six subjects with cataract were recruited and their age ranged from 61 to 87 years old (mean age, 73.5 ± 7.8 years), 20 females and 6 males. There were 13 with posterior subcapsular cataract (PSC), eight with nuclear cataract (NC) and four with cortical cataract (CC). The mean (±SD) LOS III grades for all the patients were (2.5 ± 0.8). Distance visual acuities were (mean, 0.11± 0.14) logMAR and near visual acuities were (mean, 0.23 ± 0.14) logMAR. The mean value of the spherical equivalent of the refractive error was (0.35 ± 2.30) D. The educational levels of the participants were 54% had a bachelor degree, 31% had A-level, 8% had a master degree and 8% had a PhD. All the participants were able to read English fluently.

5.1.1.1 Repeatability

The data of the reading performance metrics (RA, MRS, CPS and logRAD/logMAR ratio) for each reading chart were normally distributed (Shapiro-Wilks test $p > 0.05$). The differences of the reading performance metrics measurements between the two testing sessions were analysed for significance by means of t-test. The cut-off level for statistical significance was $P < 0.05$, two-tailed t-test. Pearson correlation coefficient and Bland-Altman agreement were performed to assess the repeatability of the reading performance measurements between the two testing sessions. Also, the coefficient of repeatability (CR) was calculated.

Table 5.1: Test-retest reliability between the first and the second testing session in cataract subjects. Pearson correlation coefficients n = 26

Reading metrics	**Colenbrander (*r* value, *p* value)**	**Radner (*r* value, *p* value)**	**MNread (*r* value, *p* value)**
Reading acuity (logRAD)	$r = 0.790$, $p = 0.000$	$r = 0.712$, $p = 0.000$	$r = 0.790$, $p = 0 .000$
Maximum reading speed (wpm)	$r = 0.703$, $p = 0.000$	$r = 0.765$, $p = 0.001$	$r = 0.860$, $p = 0.000$
Critical print size (logRAD)	$r = 0.812$, $p = 0.000$	$r = 0.694$, $p = 0.000$	$r = 0.532$, $p = 0.005$
LogRAD/LogMAR ratio (%)	$r = 0.877$, $p = 0.000$	$r = 0.882$, $p = 0.001$	$r = 0.854$, $p = 0.000$

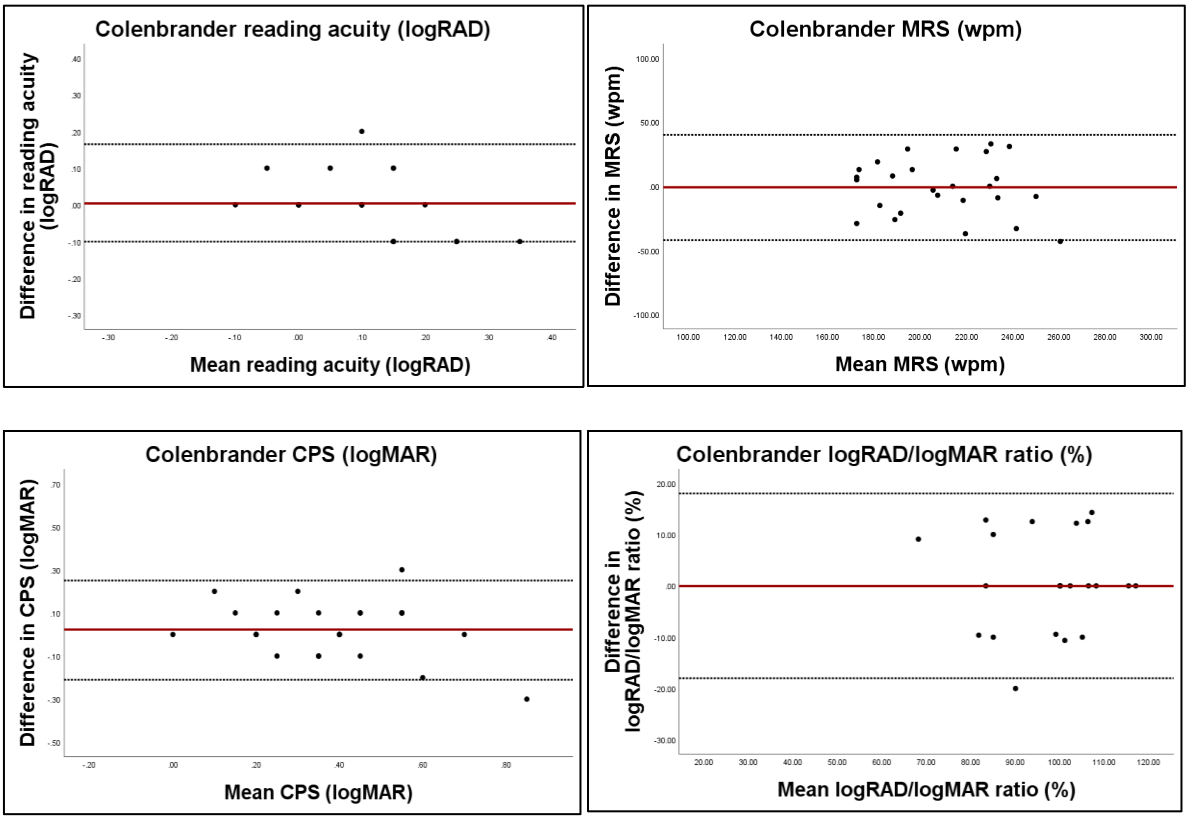
5.1.1.2 Repeatability of continuous-text reading tests (Radner, MNread and Colenbrander)

The results of the Pearson correlation coefficients analysis are shown in Table 6.1 above. As can be seen, a strong correlation was found between the two testing sessions in the three reading test charts for measuring the RA and the MRS. However, MRS results show stronger correlation in the MNread chart (*r* = 0.860) than Colenbrander (*r* = 0.703) and Radner (*r* = 0.765). For measuring CPS, moderate test-retest correlations were found in the Radner (*r* = 0.694) and MNread (*r* = 0.532) while the correlation was very strong in the Colenbrander reading chart (*r* = 0.812). The r values were above 0.8 for measuring the logRAD/logMAR ratio in the three reading charts. All the correlations were tested for statistical significance *p* value <0.05, as shown in Table 5.1.

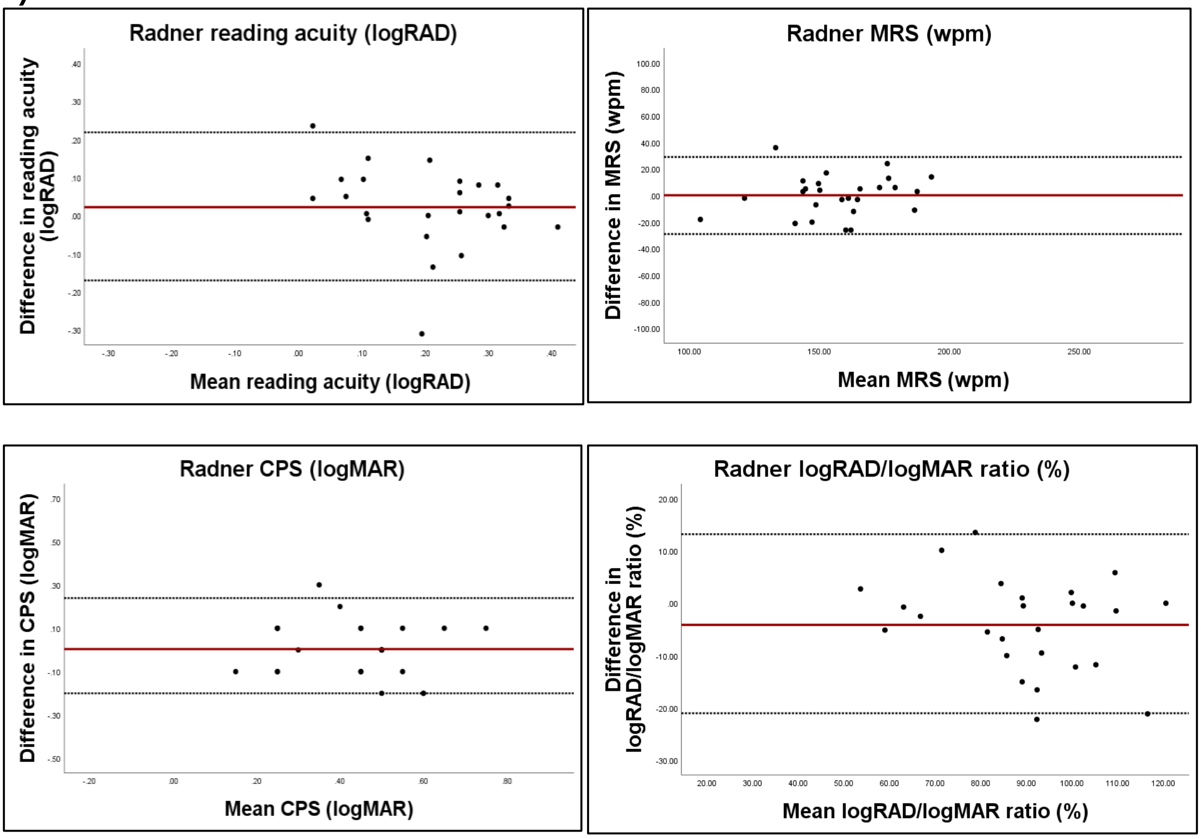
No statistically significant differences between the two sessions were found in Colenbrander and MNread for measuring all the reading performance metrics. However, in the Radner reading chart, statistically significant difference was found for measuring the logRAD/logMAR ratio. This difference was in LogRAD/logMAR ratio (%): -4.13 ± 8.83, 95% confidence interval CI: (-7.69; -0.56) *p* = 0.025. No statistically significant differences were found in the Radner charts for measuring RA, MRS and CPS.

The results of the Bland-Altman agreement analysis are shown in Figure 5.1. In the Radner and MNread reading charts, although there were differences between the two testing sessions for measuring the CPS, these were not statistically significant and low agreements were found. In the Colenbrander chart, low agreement was found for measuring the reading acuity while the agreements were high for the other metrics.

a)


Colenbrander reading acuity (logRAD)
Difference in reading acuity (logRAD)
Mean reading acuity (logRAD)
Colenbrander MRS (wpm)
Difference in MRS (wpm)
Mean MRS (wpm)
Colenbrander CPS (logMAR)
Difference in CPS (logMAR)
Mean CPS (logMAR)
Colenbrander logRAD/logMAR ratio (%)
Difference in logRAD/logMAR ratio (%)
Mean logRAD/logMAR ratio (%)

b)


Radner reading acuity (logRAD)
Difference in reading acuity (logRAD)
Mean reading acuity (logRAD)
Radner MRS (wpm)
Difference in MRS (wpm)
Mean MRS (wpm)
Radner CPS (logMAR)
Difference in CPS (logMAR)
Mean CPS (logMAR)
Radner logRAD/logMAR ratio (%)
Difference in logRAD/logMAR ratio (%)
Mean logRAD/logMAR ratio (%)

c)

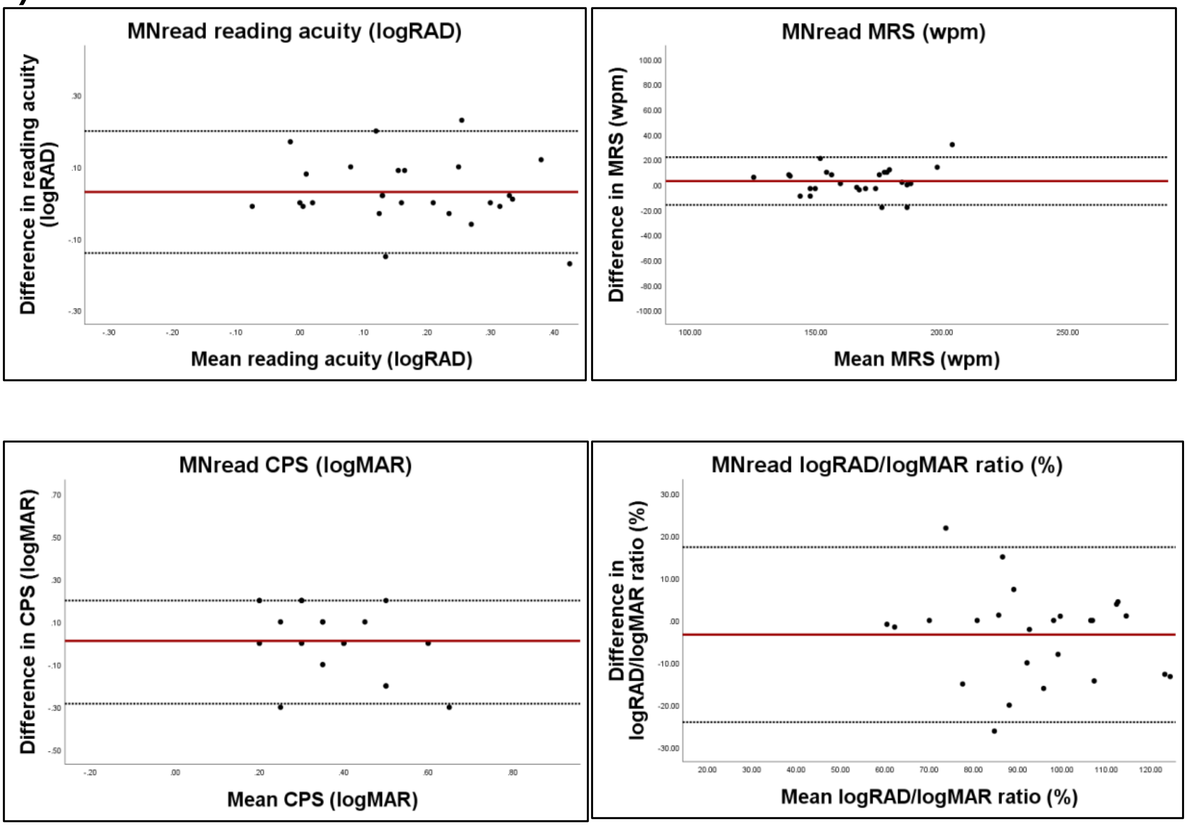


Figure 5.1: Test-retest reliability: Bland-Altman plot of the differences between the two testing sessions (first minus second) in cataract subjects. In **(a)** Bland-Altman plots for reading performance metrics in the Colenbrander chart showed high agreement between the two testing sessions for MRS, CPS and LogRAD/LogMAR ratio, but not for RA; **(b)** Bland-Altman plots for reading performance metrics in the Radner chart showed high agreements in all reading performance metrics except for CPS where the agreement was low ; **(c)** Bland-Altman plots for the MNread chart showed a low agreement for CPS, but high agreement for RA, MRS and logRAD/logMAR ratio (*n* = 26).

5.1.1.3 Repeatability of Bailey-Lovie unrelated words reading chart

The MRS (r = 0.499) and CPS (r = 0.461) measurements showed weak to moderate correlation between the testing sessions, but no statistically significant differences were found and the Bland-Altman agreements were high. Interestingly, the mean values of the CPS measurements in the two testing sessions were exactly the same and equal to 0.0477 logMAR. Strong correlations and agreements were found for measuring RA (r = 0.719) and logRAD/logMAR ratio (r = 0.840) without significant differences (Fig 5.2).

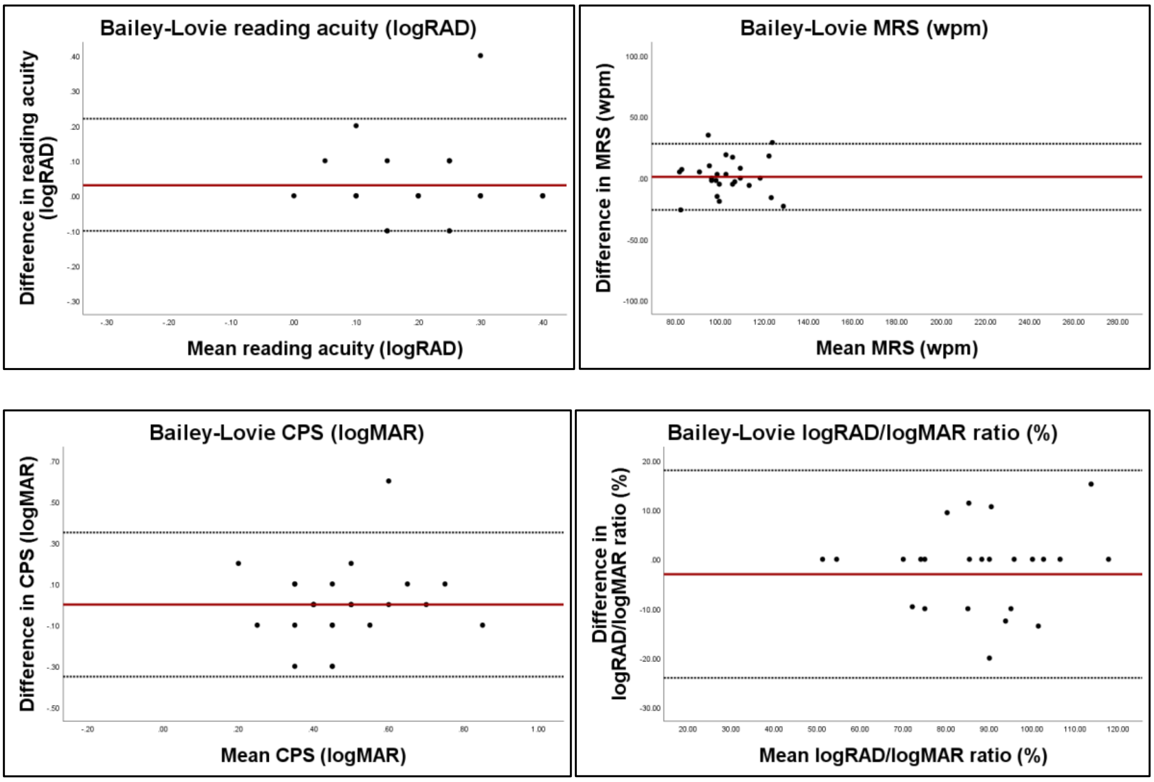


Figure 5.2: Test-retest reliability Bailey-Lovie words chart in cataract subjects: Bland-Altman plots of the differences between the two testing sessions (first minus second) for RA, MRS, CPS and LogRAD/LogMAR ratio. High agreements were found for all reading performance metrics.

5.1.1.4 Repeatability of Times New Roman chart (Ciba Vision reading card)

Again, as shown in the previous chapters with young and older age participants, the standard error of the difference for measuring the reading acuity with TNR chart was zero. Therefore, the correlation between the two testing sessions cannot be computed in people with cataract. Regarding the CPS, low Bland-Altman agreement was found. Weak correlation (r = 0.587) and agreement was found for measuring the MRS with statistically significant differences between the two sessions $p < 0.05$.

5.1.1.5 Repeatability of International Reading Speed test (IReST)

The average reading speed measurement for the three paragraphs selected (text number six, seven and eight) showed a good correlation between the two testing sessions (r = 0.795). Bland-Altman evaluated a high agreement in the repeatability of average reading speed measurements (Fig. 5.3). No statistically significant difference was found between the repeated measurements. Table 5.3 summarises the repeatability results of the reading performance metrics for the six reading charts in subjects with cataract.

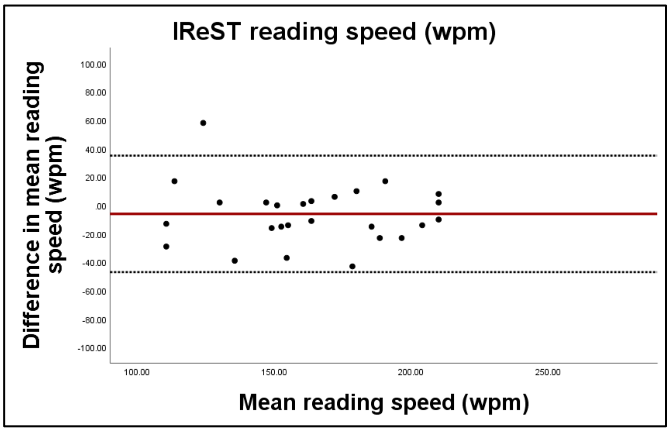


Figure 5.3: Test-retest reliability of the IReST chart in subjects with cataract: Bland-Altman plots of the differences between the two testing sessions (first minus second) for average reading speed showing a good agreement.

The coefficients of repeatability for the reading performance metrics in different reading charts are presented in Table 5.2. For MRS, Colenbrander, TNR and IReST yield more CR (less repeatable) than the other charts. MNread reading chart resulted in the lowest CR for measuring the MRS (most repeatable). TNR provided the least repeatable chart to calculate CPS in this group of patients with cataract.

Table 5.2: The coefficient of repeatability for the reading performance metrics between two testing sessions using different reading charts subjects with cataract

Reading metrics	(CR) Colenbrander	(CR) Radner	(CR) MNread	(CR) Bailey-Lovie	(CR) TNR	(CR) IReST
RA (logRAD)	0.16	0.20	0.18	0.20	-	-
MRS (wpm)	44	30	20	28	38	42
CPS (logMAR)	0.24	0.24	0.30	0.36	-	-
logRAD/logMAR ratio (%)	18.4	17.6	21.1	22.1	-	-

Table 5.3: Summary of the repeatability results of the reading performance metrics for the six reading charts in subjects with cataract

Reading chart	Reading performance metrics	Correlation coefficient (r value)	Differences in measurements (t test) *p* value	Bland-Altman agreement
Colenbrander	RA (logRAD)	0.790	0.814	[0.16; -0.10]
	MRS (wpm)	0.703	0.846	[40; -40]
	CPS (logMAR)	0.812	0.365	[0.25; -0.21]
	LogRAD/LogMAR ratio (%)	0.877	0.978	[18.0; -18.0]
Radner	RA (logRAD)	0.712	0.282	[0.21; -0.17]
	MRS (wpm)	0.765	0.949	[29; -29]
	CPS (logMAR)	0.694	0.876	[0.23; -0.20]
	LogRAD/LogMAR ratio (%)	0.882	0.025	[13.12; -21.0]
MNread	RA (logRAD)	0.790	0.111	[0.20; -0.14]
	MRS (wpm)	0.860	0.176	[0.22; -0.16]
	CPS (logMAR)	0.532	0.704	[0.20; -0.28]
	LogRAD/LogMAR ratio (%)	0.854	0.117	[17.3; -24.0]
Bailey-Lovie	RA (logRAD)	0.719	0.147	[0.22; -0.10]
	MRS (wpm)	0.499	0.626	[28; -26]
	CPS (logMAR)	0.461	1.00	[0.35; -0.35]
	LogRAD/LogMAR ratio (%)	0.840	0.173	[18.0; -24.0]
	MRS (wpm)	0.587	0.041	[30; -44]

Times New Roman	CPS (N notation)	0.717	0.346	[4; -6]
IReST	Average reading speed (wpm)	0.795	0.114	[35; -47]

5.1.2 Reliability

To assess the reliability of the reading charts in cataract subjects, Cronbach's alpha was calculated for all the reading performance metrics in each chart at the two testing sessions. For measuring the logRAD/logMAR ratio, Cronbach's alpha yielded an excellent reliability in Colenbrander, Radner, MNread and Bailey-Lovie ($\alpha > 0.9$). In Colenbrander and Radner reading charts, high reliability results were found for measuring the RA, MRS and CPS ($\alpha > 0.8$). MNread chart noticeably revealed higher reliability result for measuring the MRS ($\alpha = 0.925$) compared to the other charts. However, relatively weak correlation was found for measuring the CPS ($\alpha = 0.694$). In the Bailey-Lovie reading chart, high reliability result was found for measuring the RA ($\alpha = 0.836$), but not for MRS ($\alpha = 0.661$) and CPS ($\alpha = 0.631$). In the TNR chart, the Cronbach's alpha could not be computed for RA. However, good reliability was found for measuring CPS ($\alpha = 0.835$) followed by MRS ($\alpha = 0.740$). In the IReST reading chart, good Cronbach's alpha was found for measuring the average reading speed in cataract subjects ($\alpha = 0.886$). Table 5.5 summarises the reliability results of the reading charts.

5.1.3 Comparison of the reading errors between the charts

Table 5.4: Number of reading errors per reading chart in subjects with cataract

Test session	**Colenbrander**	**Radner**	**MNread**	**Bailey-Lovie**	**Times New Roman**	**IReST**
First (mean± SD)	6.15± 3.81	4.92± 4.00	2.31± 2.68	2.96± 2.42	6.08± 5.16	4.77± 4.34
Second (mean± SD)	5.15± 3.17	4.08± 3.21	2.15± 1.84	2.62± 2.11	4.73± 4.33	5.38± 5.81

Table 5.5: The Cronbach's alpha coefficient results for evaluating the reliability of the reading chart in presbyopic subjects.

Reading chart	Reading performance metrics	Cronbach's alpha	Reading chart	Reading performance metrics	Cronbach's alpha
Colenbrander	RA (logRAD)	0.883	**Bailey-Lovie**	RA (logRAD)	0.836
	MRS (wpm)	0.826		MRS (wpm)	0.666
	CPS (logMAR)	0.897		CPS (logMAR)	0.631
	LogRAD/LogMAR ratio (%)	0.934		LogRAD/LogMAR ratio (%)	0.913
Radner	RA (logRAD)	0.832	**TNR**	MRS (wpm)	0.740
	MRS (wpm)	0.867		CPS (N notation)	0.835
	CPS (logMAR)	0.819			
	LogRAD/LogMAR ratio (%)	0.937	**IReST**	Average reading speed (wpm)	0.886
MNread	RA (logRAD)	0.883			
	MRS (wpm)	0.925			
	CPS (logMAR)	0.694			
	LogRAD/LogMAR ratio (%)	0.921			

The data of the reading errors were not normally distributed (Shapiro-Wilks <0.05). A Wilcoxon signed-rank test was conducted to determine the median differences in each reading chart between the first and second sessions as well as the differences of the number of reading errors between the charts in the first testing session. No statistically significant differences were found in the number of reading errors in all reading test charts between the first and the second testing sessions (p <0.05). Comparing the reading errors between the reading charts in the first visit showed that the Bailey-Lovie word reading chart resulted in fewer reading errors compared to all other reading charts except for the MNread chart, the differences between them were not statistically significant. These median differences were as follows: between Bailey-Lovie and Colenbrander (2.84 words) p = 0.008, Bailey-Lovie and Radner (2.42 words) p = 0.015, Bailey-Lovie and TNR (2.93 words) p = 0.003 and between Bailey-Lovie and IReST (2.08 words) p = 0.038. On the other hand, the Colenbrander reading chart resulted in significantly more reading errors in cataract participants compared to MNread and IReST and these differences were 2.84 words and 2.00 words, respectively $p < 0.05$. No significant differences were found in a number of reading errors between all the other reading charts.

The coefficient of determination r^2

A multiple regression analysis was run to determine to what extent in percentage the types and grades of the cataract could influence the reading performance results in each reading chart. Table 6.6 shows that the Bailey-Lovie reading chart shows the minimal dependence on the cataract type and grade, as the coefficient of determination was small and less than 6%. In this analysis, small coefficient of determination r^2 means that a high percentage of the data variance is accounted by the reading test itself. For the variable RA, MRS, CPS and logRAD/logMAR ratio, the cataract type and grades accounted for 5.5-10.8% of the variance in the Colenbrander reading chart, while r^2 was higher in Radner and MNread and accounted for 2.9- 24% and 0.9- 30% of the variability, respectively (Table 5.5). The different cataract types markedly influenced the results of the MRS in the IReST (28.6%) and TNR charts (35.9%).

Table 5.6: The coefficient of determination r^2 in percentage: the influence of the cataract type and grade on the variance of the reading performance metrics (a lower percentage suggests less influence from the cataract).

Variables	**Colenbrander**	**Radner**	**MNread**	**Bailey-Lovie**	**IReST**	**TNR**
RA (logRAD)	10.8%	20.7%	23.3%	5.4%		
MRS (wpm)	8.6%	10.9%	15.2%	5.2%	28.6%	35.9%
CPS (logMAR)	2.4%	24%	30.2%	2.4%		
logRAD/logMAR ratio	5.5%	2.9%	0.9%	0.9%		

5.2 Discussion

5.2.1 Repeatability and inter-chart reliability of the reading test charts for measuring the reading performance metrics in people with cataract

The inclusion of a reading test in clinical examination of patients with cataract is necessary to facilitate the clinical diagnoses or planning a surgical intervention. As shown in previous chapters the reading performance of healthy individuals varies with the test type and this may be exaggerated with ocular pathology such as cataract.

The results of the reliability and repeatability of the logRAD/logMAR ratio are more important to considerable in cataract subjects than normally sighted subjects. This is because the reading acuity and distance visual acuity can differ significantly in some eye diseases, including cataract (Stifter et al., 2005c). Colenbrander, Radner, MNread and Bailey-Lovie reading charts revealed a good test-retest and inter-chart reliability for measuring logRAD/logMAR ratio.

IReST reading chart provides good repeatability and reliability results in patients with cataract. However, the coefficient of repeatability for measuring the reading speed was higher than that in the Bailey-Lovie, MNread and Radner. In addition, the Cronbach's alpha value was less than that found in the previous chapters with young and older participants. The most likely cause is that the IReST reading test consists of one print size 1.0 M (newspaper-size text) and it has been proven that the reading speed of the small text is significantly reduced by cataract. However, when the text is enlarged enough, the reading speed of cataract patients may not be affected and match the normal level (Elliott et al., 2000). It has been suggested that longer text paragraphs, such as in the IReST chart, are more reliable and represent the real-life reading performance (Trauzettel-Klosinski et al., 2012). However, for subjects with cataract, it is important to evaluate the spot reading for daily activities, such as reading email or price tags. Therefore, the short sentences reading test may work best in this case. These findings corroborate the finding of the high coefficient of determination r^2 (28.6%), which indicates that the variation of the MRS measurements in the IReST reading chart depend on the cataract type and grades rather than the chart itself. This result matches those observed in the TNR chart.

Reading acuity is one of the most important metrics for evaluating the reading performance in subjects with cataract. In this study, the TNR chart resulted in unreliable results for measuring the reading acuity. The CPS measurements in the TNR chart showed low Bland-Altman agreement. These results further support the idea that the TNR chart is not a good choice to measure the reading performance. Although there is no statistically significant difference for measuring the RA in Colenbrander reading chart, the agreement between the two visits was low.

The results show that the MNread chart had a good repeatability for RA, MRS, logRAD/logMAR ratio and scored highest reliability results for measuring these reading metrics compared to the other reading charts. However, the CPS had lower reliability and agreement between the two testing sessions. These findings are in agreement with those of Virgili et al. (2004a) and Subramanian and Pardhan (2009). In addition to the factors that explained the poorer repeatability of CPS results in the previous chapters, another reason is that the CPS measures in much larger steps than the RA, (0.1 logMAR vs. 0.01 logMAR) for the MNread chart and (0.1 logMAR vs. 0.005 logMAR) Radner chart. This outcome is contrary to that found in the Colenbrander reading chart as the CPS showed a good repeatability and reliability results. As previously stated that the word value in the Colenbrander chart is not as clear as in the Radner and MNread charts, it seems possible that the same steps used to calculate the CPS and RA (0.1 logMAR vs. 0.1 logMAR) contributed to reduce the variance for measuring the CPS in the Colenbrander chart for this group of subjects.

The coefficient of the repeatability of the Radner reading chart in this study is in markedly higher than that found in the study of Maaijwee et al. (2008b) which assessed the repeatability of the Radner reading chart in Dutch languages in older subjects affected by macular disease. Their coefficients of repeatability were 0.036 logRAD for RA, 8.2 wpm for MRS and 0.06 logMAR for CPS compared with our results of 0.20 logRAD for RA, 30 wpm for MRS and 0.24 logMAR for CPS. The difference in the repeatability of the RA might be due to including different stages of macular diseases in their study, which may have influenced the reading variables. Also, they performed the reading test at distance of 28cm rather than 40cm.

The Bailey-Lovie unrelated word reading chart is highly standardised and resulted in high Bland-Altman agreements for all the reading performance metrics between the two sessions without any statistically significant differences. Also, the Bailey-Lovie chart showed only a minimal dependence on the subject's type of cataract or grades on the results of the reading performance metrics as the coefficients of determination were small. As stated earlier, the print size should be the only parameter that affects the reading performance measurements throughout the chart. It can, therefore, be assumed that prediction of the next word in the reading sentence is an important element enabling high test-retest and inter-chart reliability of the reading test. These findings have been justified for subjects with maculopathy; however, they used unrelated word SKread test (MacKeben et al., 2015).

5.2.2 Comparison of the reading errors between the charts

It was questioned in Chapter Four whether the presence of the eye disease contributes to more reading errors with an unrelated word reading chart compared to the continuous text reading charts, as found in the study of Nair et al. (2006). The results showed that the cataract subjects made significantly fewer reading errors with the Bailey-Lovie word chart compared to the other reading charts. This result is similar to those reported in the previous chapters with young and older normally sighted participants. Thus, the current study does not support the results of Nair et al. (2006) , because the serious weakness of their study was that not all the participants were native English speakers, just 79% of them were fluent in English, and this probably would affect the result of the reading errors made. The study presented in this chapter and the study by Nair et al. (2006) both had used subjects with similar levels of education, as it has been shown that the educational level had an important influence on the reading performance results (Munch et al., 2016).

5.3 Conclusion

In the optometric practice , it is becoming necessary to use reading tests that reproducibly measure the impact of the ocular disease, such as cataract, on the patient's quality-of -life and to assist practitioners with patient management . The results of this study suggest that the Bailey-Lovie word reading chart and MNread chart provide reliable and reproducible results of the reading performance metrics for individuals with cataract. The findings of this investigation complement those of earlier studies (Chapter 3 and 4) for young and older normally sighted individuals.

Chapter 6

Comparing the Reading Metrics in Presbyopia Correction with Constriction of the Pupil

6 Introduction

The traditional methods to help people with presbyopia are spectacles and contact lenses, ,monovision or multifocal. In addition, there are surgical corrections of the crystalline lens, the cornea (lasers or inlays) and the sclera. Surgical approaches include intraocular lenses, laser ablation and scleral surgery.

One of the surgical techniques that employs the use of a corneal inlay, commercially known as Kamra (Corneal Inlay ACI7000, AcuFocus,Inc), is based on the principle of increased depth of focus by introducing an artificial reduced diameter pupil, thus improving the near visual acuity (Naroo and Bilkhu, 2016). Several approaches have been attempted with ophthalmic drops that induce pupil miosis to achieve an increase in depth of focus without surgery. Such an approach is not new and carbachol with brimonidine (Abdelkader, 2015), pilocarpine hydrochloride (Gwon and WoldeMussie, 2002) , echothiophate (Nolan, 2001), and brimonidine (Kesler et al., 2004) are amongst the ophthalmic drops that have been studied. The major problem is that studies to date have tended to focus on near visual acuity outcomes (single letter optotype) rather than reading performance outcomes. However, a standardized measurement of the reading performance by using a valid reading chart is quite important to allow accurately and efficiently evaluating the efficacy of presbyopic correction. The previous chapters evaluated the reliability and repeatability of different reading charts in three groups of subjects (young normally sighted, presbyopia and cataract). In this chapter another small group of presbyopic people added to see to which extent using a reliable reading chart and performing the full reading assessment would help in the clinical decision in the treatment of presbyopia.

Pilocarpine, a parasympathomimetic, is known to induce miosis and to decrease the intraocular pressure (Krill and Newell, 1964). Pilocarpine eye drops are commercially available in concentrations of 1%, 2% and 4%. The advantage of using the low dosage is to minimize any adverse effect such as headache, as it is well known from glaucoma patient experience that pilocarpine causes a dull headache on instillation of the drop (Bartlett and Jaanus, 2007).

The aim of this study was to perform standardized measurement of near visual ability to allow evaluating the efficacy of using low concentration pilocarpine drops (0.5%) to create optical miosis.

6.1 Subjects and methods

In the present study 10 presbyopic participants were recruited. The age range of subjects was from 49 to 69 years (mean, 60.2 ± 6.5). A small sample was chosen in this pilot study to address the effect of pilocarpine drops on reading performance metrics. Exclusion criteria include those with ocular diseases or those who have undergone intraocular surgery, use of medication which may influence cognitive ability and the presence of binocular vision anomalies and reading disorder (dyslexia). Subjects with best corrected visual acuity at distance and near of worse than 0.1 logMAR were excluded. Also, the exclusion criteria included subjects with iritis or pupillary block glaucoma, diabetes or hypertension with clinical evidence of retinal pathology, and those with age-related macular degeneration. Ethical approval was obtained from Aston University Ethical Committee and all the subjects gave their consent to participate in the study (see Appendix 4).

General health history and ocular history were recorded. Ocular surface slit lamp examinations were performed on all participants to ensure the absence of any inflammation before pilocarpine drop instillation. Distance visual acuity was measured monocularly for each eye and binocularly with and without distance correction using a standardised Early Treatment Diabetic Retinopathy Study (ETDRS) chart displayed by software CSO vision chart v2.6.0 on an LCD screen at a distance of 4m. Near visual acuity was measured monocularly for each eye and binocularly with and without near correction using an ETDRS near vision chart (Precision Vision) at 40cm. Refraction was performed for all participants.

The reading test chart used in this study was the Radner reading chart. To ensure fluency in the English language, which may affect reading performance results, all the participants were native English speakers.

Although the MNread chart resulted in better repeatability and reliability results in presbyopic subjects (Chapter 4), the Radner chart was chosen to avoid repetitions and memorization during the test, as it has three different versions whilst MNread chart has only one.

The Radner reading test charts were presented on a reading stand at a distance of 40cm, determined by using a centimetre ruler. To ensure a constant viewing distance, a headrest for the forehead was used. The luminance from all the reading test charts was 120 cd/m².

Each reading chart was covered by a piece of paper and the participants were asked to read the sentences as soon as the examiner removed the cover. The participants were instructed to read the sentences one after the other, loudly and accurately, as quickly as possible without correcting themselves when they made any errors and without repeating the words or stopping in-between. Reading performance metrics were measured monocularly and binocularly with and without near correction.

All testing procedures were monitored using audio recording to measure the reading time for each sentence and to record any reading errors. Reading performance metrics measurements were carried out after the testing session, from the participants' audio recordings. GoldWave Inc (St. John's, Newfoundland, Canada) software was used for playing the audio recordings.

Following the visits, the reading performance metrics outcomes were described as follows: reading acuity (logRAD score), average reading speed (ARS), maximum reading speed (MRS), and critical print size (CPS).

All the subjects then received 0.5% pilocarpine eye drop in their non-dominant eye. In this experiment, the dominant eye was determined by the eye used for sighting when the subject looked at a distant target through a ring hole held in both hands and with both eyes open.

A NeurOptics VIP-300 pupilometer (NeurOptics, Inc.) was used to measure pupil size before and after instillation of pilocarpine according to the company instructions. The pupilometer is a hand-held infrared device which uses monocular occlusion and objectively measures the maximum pupil diameter before constriction and minimum pupil diameter at peak constriction at different background illumination. Mean and standard deviation of the maximum and minimum pupil diameters were calculated.

Intraocular pressure (IOP) was recorded for each eye before and after instillation of pilocarpine using a non-contact tonometer (Reichert® 7CR Auto Tonometer). This machine uses an air-puff and shows the results of three recordings and the average for each eye. The average value was recorded. It has been found that the non-contact tonometer provides results comparable to those of the gold standard (Goldman Applanation Tonometer) (Yilmaz et al., 2014).

Defocus curves were evaluated binocularly and for the non-dominant eye, ranging from +1.50D to -5.00D with randomized lenses and letter sequences to decrease the effect of memory (Gupta et al., 2007).

The participants were completed the Near Activity Visual Questionnaire (NAVQ) before pilocarpine instillation to assess near vison ability with their current near correction, and after five hours of pilocarpine instillation to assess their near vision ability with the effect of pilocarpine only and without their habitual near correction. The participants were also asked to report any physiological or ocular reactions that they experienced after pilocarpine instillation (see Appendix 5).

One hour after pilocarpine instillation the participants returned. Their near and distance visual acuities, reading performance metrics, pupil size, IOP, defocus curve and NAVQ questionnaire were evaluated before and after pilocarpine instillation, at the same room illumination.

6.1.1 Statistical analysis

The data were normally distributed as assessed by Shapiro-Wilk's test ($p > 0.05$). A paired samples t-test was used to determine whether there was a statistically significant mean difference in the measurements before and after pilocarpine instillation. The cut-off level for statistical significance was $p<0.05$. The statistical analysis was performed using Statistical Package for Social Sciences (SPSS) version 23.

6.2 Results

Presbyopic subjects (one male and nine females) with an age range of 49 to 69 years (mean, 60.2 ± 6.5) were recruited in this study. There were four with right eye dominance and six with left eye dominance.

6.2.1 Non-dominant eye ocular measurements (the eye received 0.5% pilocarpine drop)

In the non-dominant eye which received 0.5% pilocarpine drops, the maximum pupil size (before constriction) decreased significantly from 3.9 ± 0.34 mm before pilocarpine instillation to 2.6 ± 0.42 mm at 1 hour after pilocarpine instillation ($p < 0.0001$). Maximum pupil diameter had a significant decrease of 1.4 mm (95% CI, 0.99 to 1.7) mm. Also, the minimum pupil size (peak of constriction) showed a statistically significant decrease from 2.3 ± 0.35mm to 1.9 ± 0.21 mm at 1hr after pilocarpine instillation ($p < 0.001$). Minimum pupil size decreased by 0.38 mm (95% CI, 0.19 to 0.57) mm.

No statistically significant differences were found in the IOP measurement in the non-dominant eye before and after 1hr of pilocarpine instillation ($p > 0.05$).

The mean value of the spherical equivalent of the refractive error was (-0.21± 1.89) in the non-dominant eye. This spherical equivalent was exactly the same after 1hr of instillation and no statistically significant difference was found (p =0.36).

Mean change in Distance Uncorrected Visual Acuity (DUCVA) and Distance Best Corrected Visual Acuity (DBCVA) measured with ETDRS at 4 m have been compared before and after pilocarpine instillation. Also, Near Uncorrected Visual Acuity (NUCVA) and Near Best Corrected Visual Acuity (NBCVA) measured with 40 cm ETDRS determined whether there was a statistically significant mean difference before and after 1 hour of pilocarpine instillation.

Statistically significant improvements were found in DUCVA (p =0.007) and NUCVA (p= 0.018). DUCVA improved significantly from 0.19 ± 0.24 logMAR before pilocarpine instillation to 0.08 ± 0.19 logMAR after 1 hour instillation. While NUCVA improved significantly from 0.36 ± 0.14 logMAR to 0.24 ± 0.11 logMAR. No statistically significant differences were found in the best corrected visual acuities before and after pilocarpine instillation for distance or near.

No statistically significant differences or improvements were found in the reading performance metrics (logRAD score, ARS, MRS and CPS) measured with near correction before and after pilocarpine instillation. logRAD score measured without near correction statistically improved from 0.4±0.14 logRAD to 0.3 ± 0.11 logRAD at 1 hour of pilocarpine instillation(p = 0.021). Although the logRAD score measurements without near correction significantly improved after instillation, this improvement was less than one line on the chart, which may be clinically insignificant. The other reading performance metrics measured without near correction (ARS, MRS, and CPS) were not statistically significantly improved after pilocarpine instillation. Table 6.1 summarizes the mean change in the reading performance metrics measurements in the non-dominant eye.

Table 6.1: Mean change in reading performance metrics measurements in the non-dominant eye before and after one hour of pilocarpine instillation

Reading performance metrics	*Mean before pilocarpine instillation*	*Mean after 1 hour of pilocarpine instillation*	*t-test (p- value)*
Without near correction			
logRAD score	0.4	0.3	0.021
ARS (wpm)	144	138	0.279
MRS (wpm)	174	174	0.952
CPS (logMAR)	0.6	0.5	0.061
With near correction			
logRAD sore	0.01	-0.01	0.364
ARS (wpm)	165	161	0.454
MRS (wpm)	181	176	0.354
CPS (logMAR)	0.2	0.2	0.591

Figure 6.1 shows the mean monocular defocus curve difference before and after pilocarpine instillation. The monocular defocus curve showed a significant improvement in visual acuities at the vengeance corresponding to the very far distance vision (+1.50D, +1.00D, +0.5D; $P < 0.05$) and for the very near distance vision (lens power from -3.50D to -5.00D; $p < 0.05$). No statistically significant differences were found for the distance, intermediate and near vision (lens power from 0D to -3.00 D; $P > 0.05$).

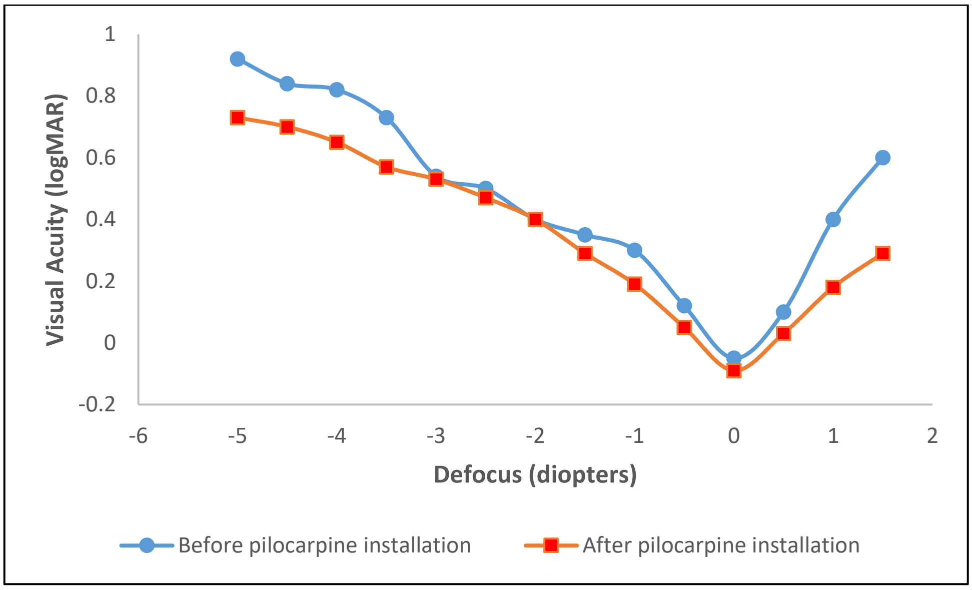


Figure 6.1: Mean monocular defocus curve before and after pilocarpine instillation

Pupil size returned to normal after 8 hours instillation there was no statistical difference between pupil size before pilocarpine instillation and after 8 hours instillation.

6.2.2 Binocular measurements

Binocular visual acuities and reading performance were performed to assess the improvement in vision ability before and after 1 hour of pilocarpine instillation. No statistically significant differences were found in the BDUVA (*p* =1.77). A very slight improvement in the binocular distance DBCVA -0.1±0.08 logMAR before pilocarpine instillation to -0.14±0.05 logMAR at 1 hour of instillation (*p*= 0.03). However, this improvement was very small, at 0.04 logMAR. Although, for the binocular uncorrected near visual acuity no statistical differences was found (*p*=0.056), the difference in the mean was around 0.1 logMAR which represents one line of improvement (mean difference, 0.09 ± 0.12). No statistical difference in the binocular near best corrected visual acuity before and after pilocarpine instillation (*p* =0.11).

Table 6.2 shows the mean change in reading performance metrics measured binocularly with and without near correction before and after 1 hour of pilocarpine instillation. The results showed no effect of pilocarpine drop on reading performance metrics measured binocularly with and without near correction.

Table 6.2: Mean change in reading performance metrics measured binocularly in the eye before and after one hour of pilocarpine instillation

Reading performance metrics	*Mean before pilocarpine instillation*	*Mean after 1 hour of pilocarpine instillation*	*t-test (p- value)*
Without near correction			
logRAD score	0.42	0.25	0.780
ARS (wpm)	159	156	0.625
MRS (wpm)	181	186	0.452
CPS (logMAR)	0.56	0.48	0.053
With near correction			
logRAD sore	-0.05	-0.02	0.083
ARS (wpm)	165	168	0.977
MRS (wpm)	182	184	0.704
CPS (logMAR)	0.2	0.2	1.00

Figure 6.2 shows the mean binocular defocus curve difference before and after pilocarpine instillation. No statistically significant differences were found in visual acuity in all lens powers from +1.5D to -5.0 D after pilocarpine drop instillation ($p > 0.05$).

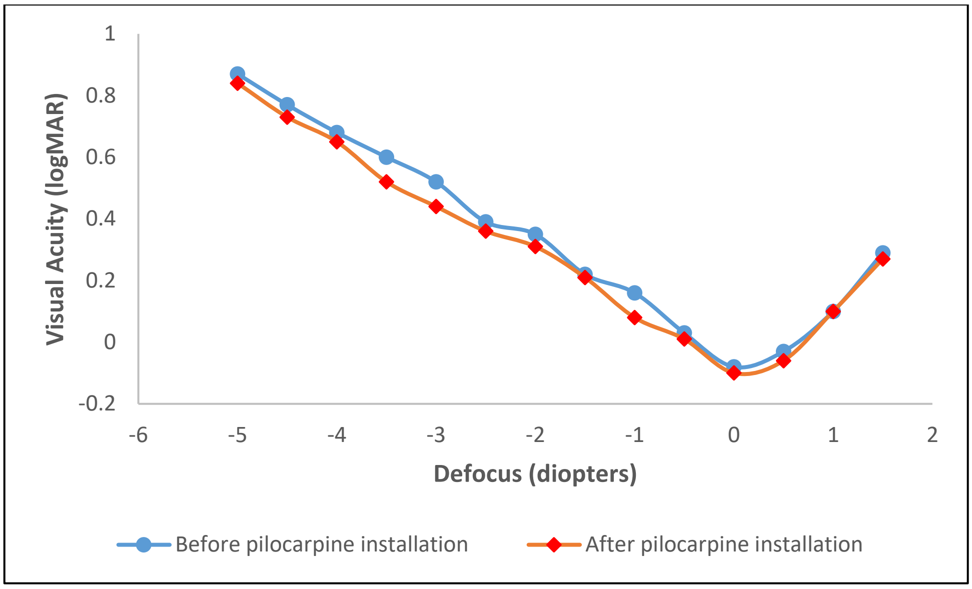


Figure 6.2: Mean binocular defocus curve before and after pilocarpine instillation

The participants' subjective perception of their near vision ability was determined by administering the Near Activity Visual Questionnaire (NAVQ) which was scaled to the Rach score (100 = extreme difficulty, 0 = no difficulty). The mean subjective perception of the participants with their habitual near correction before pilocarpine drop instillation was 27.55 ± 10.79 log-odd units. However, the participants' satisfaction about their near vision ability after 5 hours of pilocarpine instillation was slightly worse (32 ± 27.43 log-odd units). This differences in NAVQ scores before and after pilocarpine instillation were not statistically significant (p = 0.570).

6.3 Discussion

This study assessed the efficacy of low concentration pilocarpine drops for the improvement of near vision in presbyopia by comparing the reading performance metrics in addition to the visual acuities before and after 1 hour of instillation using standardized type of reading chart (Radner). The treatment of only one eye using the low dosage of pilocarpine has several advantages; reduced symptoms of headache and, no ocular complications were observed in any subjects during or after the examination and it does not blur the distance vision.

Regarding the amount of pupil size reduction, 0.5% pilocarpine reduced the pupil size statistically significant from 3.9 ± 0.34mm before pilocarpine instillation to 2.6 ± 0.42mm at 1 hour after pilocarpine instillation. Comparing to other methods of presbyopia correction that used the same principle of small pupil size, 0.5% pilocarpine reduced the pupil size to a lesser extent. Intracorneal inlay and pinhole contact lens resulted in reduced the pupil size to 1.6mm

(Yilmaz et al., 2008, García-Lázaro et al., 2012). Also, 4% pilocarpine reduced pupil size to 1.8mm (Zaczek and Zetterström, 1998). It has been suggested that 1.6mm pupil size provides good depth of field and decreases the effect of diffraction on the visual acuity (Yilmaz et al., 2008). This is because reducing the pupil size will also influence the amount of light transmitted to the eye.

Defocus curve test is an effective way to assess the clear range of vision (Gupta et al., 2009a). In the non-dominant eye, the mean visual acuity measurement as a function of defocus curve showed significantly better visual acuities after pilocarpine instillation for very far distances between 200cm to 67cm and for very near distances between 33cm to 20cm. No improvement was found in the near and intermediate distances. For binocular vision, no improvement was found in all distances. Therefore, pilocarpine does not seem to improve the functional intermediate vision. Intermediate vision can be demanded in presbyopic people because of current lifestyles, for example using a computer. Therefore, any presbyopia solution should provide a wide range of vision.

In the present study, binocular DBVA and the monocular DUVA were improved, which gives an advantage of pilocarpine drops to maintain good distance visual acuities. After pilocarpine instillation, binocular and monocular uncorrected near visual acuities results measured with ETDRS showed an improvement for about one line. However, the near VA disagreed with reading acuity and cannot predict the reading performance in presbyopic subjects. The full reading assessment was performed using the Radner reading chart. The findings provide further support and showed that, despite the improvement of the near VA, the reading performance metrics (RA, MRS and CPS) showed no statistically significant differences after pilocarpine drops administration. The importance of using a standardized reading chart can be further explained by the NAVQ results as the subjects' satisfaction about reading different material was worse despite the improvement in the near VA.

A number of studies, such as Nolan (2001), Gwon and WoldeMussie (2002), Kesler et al. (2004) and Abdelkader (2015), which evaluated the effect of the ophthalmic drops to treat presbyopia used the near ETDRS chart or Jaeger chart to assess the near vision and failed to assess the reading ability. The Jaeger chart has been criticized for the lack of consistency and failure to follow the meaningful size progression (Rubin, 2013).

There are possible limitations of this study, including the small sample size and the study would be better performed over a longer follow-up to properly address the effect.

6.4 Conclusion

Improving depth of focus by reducing pupil diameter using 0.5% pilocarpine in the non-dominant eye showed an improvement in near vision acuity measured with EDRS chart (isolated letters) in presbyopic subjects while keeping good distance visual acuity. However, an improvement of the reading performance metrics with decreasing pupil diameter was not found. There is, therefore, a definite need for using a standardized reading test chart to evaluate the effectiveness of presbyopia treatment and to help for accurate clinical decision.

Chapter 7

The accuracy and the validity of the mobile app tablet-based reading tests

7 Introduction

The increased use of personal electronic devices is creating novel ways to improve health practices. The mobile phone technology applications, ‘apps’, in tablet-based approaches, in particular the iPad tablet, have been evaluated as alternative to original paper-based charts for testing the visual functions, such as visual acuity and contrast sensitivity measurements tablet apps (Livingstone et al., 2016, Kingsnorth et al., 2016). However, only two apps are available to measure the reading performance metrics in a way similar to traditional paper-based reading performance charts. These reading test apps are the Radner app, developed by Kingsnorth and Wolffsohn (2015), and the MNread app, developed by (Calabrese et al., 2014). Only one study for each app was available in the literature and was conducted by the developers of the reading apps and never evaluated for test-retest reliability. Therefore, these reading apps are in need of further robust independent evaluation.

With paper-based reading charts the examiner often uses a stopwatch to measure the reading time and record the reading errors manually at the same time, and this is a time-consuming process. In addition, the CPS is determined subjectively by the examiners by plotting the reading performance data graphically and this could increase the variability, as it has been shown in the previous chapters that there is presence of some noisy data around the CPS when comparing it to other studies. Turning to automated estimation of the reading metrics may overcome these issues and standardise the results and this becomes possible with the use of new technology to assess reading performance.

This chapter has investigated the accuracy of the MNread and Radner tablet-based applications with the data obtained from young normally sighted, presbyopic and cataract participants. Firstly, this study evaluated the test-retest and inter-chart reliability of the MNread and Radner apps. Secondly, this study compared the outcome of the reading performance metrics between the two tablet-based applications. Finally, it compared the reading performance of the participants on the MNread and Radner tablet-based apps with the paper-based charts of both.

7.1 Methods

The participants performed two tablet-based reading tests (MNread and Radner) in the two testing sessions, which were two to three weeks apart. Twenty-nine young normally sighted participants (mean age, 20.4 ± 3.7 years), 29 presbyopic participants (mean age, 61.5 ± 8.8 years) and 26 participants affected with cataract (mean age, 73.5 ± 7.8 years) were recruited. Exclusion criteria include subjects for whom English is not their first language, those with ocular diseases other than the cataract in cataract group of subjects, educational level less than A-level, use of medication that may influence the cognitive ability and presence of binocular vision anomalies and reading disorder (dyslexia). All the tests were performed monocularly with the most up-to-date near corrections, if needed. Data were collected with informed consent and the study was conducted in accordance with the Helsinki declarations.

7.1.1 Procedure

Four charts were used for this study, Radner and MNread paper-based charts and Radner and MNread tablet-based applications. The Radner reading speed test app was created by Kingsnorth and Wolffsohn (2015) and licenced to Aston University EyeTech. It was created for iPad 3 using Apple's X-code Starkit Developer eXtension in the Objective-C programming language.

As in the Radner paper-based chart, each sentence in the app consists of 14 words and are arranged in a logarithmic progression. The range of the text sizes in the Radner app was from 0.9 to -0.2 logMAR, while in the Radner paper chart the range was from 1.1 to -0.2 logMAR. Once the test started, the Radner app would present the reading sentences in 0.1 logMAR steps once at a time (Fig.7.1). Two buttons appeared at the bottom of the screen "Read" and "Cannot Read". When the participant reaches the smallest text size that can be read, he/she has to press "Cannot Read" button to end the test and after that reading acuity is recorded simultaneously. For young and older subjects, the test was managed by themselves after being instructed and doing one trial test before the actual test began. The self-administration was employed to help the participant to engage and motivate during testing. On the other hand, for the cataract group of subjects, the test was managed by the examiner to avoid any delay in the response time. The Radner app simultaneously recorded the voice of the subject and used the stopwatch to measure the reading time for each sentence from when the sentence was presented until the "Read" button had been pressed.

The developed version of the MNread app was purchased from Apple iTunes store and implemented on the iPad 3 tablet. The app was developed by Calabrese et al. (2014) in the University of Minnesota with retinal display, resolution 264 pixels per inch. Like the MNread paper-based chart, the sentences are presented in three lines and arranged in a logarithmic progression. However, the number of the sentences and the range of print sizes are reduced in the app version (14 sentences compared to 19 sentences in the paper chart). In the MNread app the print sizes range from 1.2 to -0.1 logMAR, while in the paper chart they range from 1.3 to -0.5 logMAR. The MNread app is available in different languages (English, French, Greek, Italian, Portuguese, Spanish and Turkish). Since the MNread app is not suitable for self-administration and should be conducted by the examiner, an external keyboard was connected to the iPad via Bluetooth to manage the MNread app test by the examiner and to avoid subject's distraction during the test by the finger tab on the screen (Fig 7.1). Each sentence was presented once at a time; once the sentence appeared on the screen it was accompanied by a beep and the subjects were instructed to read the sentences as fast as they could directly after they heard the beep . After each sentence was read, a score screen for the reading errors was displayed to the examiner to enter the number of errors made by using the external keyboard, see Figure 7.2. The test stopped when the participants could no longer read any words. The MNread app then displayed the corresponding graphical plot of the reading speed as a function of the print size. Figure 7.2 shows the testing sequence in the MNread app.

The Radner app contains two versions of charts while the MNread app contains five versions; different charts were used in the two testing sessions to avoid memorisation.

For both iPad apps, the screen luminance was set to 200cd/m2. The participants were positioned at 40cm from the iPad screen or the paper chart. To ensure constant viewing distance, a headrest for the forehead was used.

For the paper-based charts the procedure was performed as described in Chapter 3 using the audio recording to measure the reading performance metrics. Returning briefly to the scoring methods used to calculate the reading performance metrics in the paper-based charts, these were as follows:

- RA in logRAD was determined as the smallest print size that could be read, taking the reading errors into account (0.005 logMAR for incorrectly read syllables in Radner and 0.01 for incorrectly read word in MNread)
- CPS in logMAR was determined as the steepest decline slope on the graph of plotted reading time in wpm per print size

- MRS in wpm was the average reading speed of the sentences corresponding to and larger than CPS.

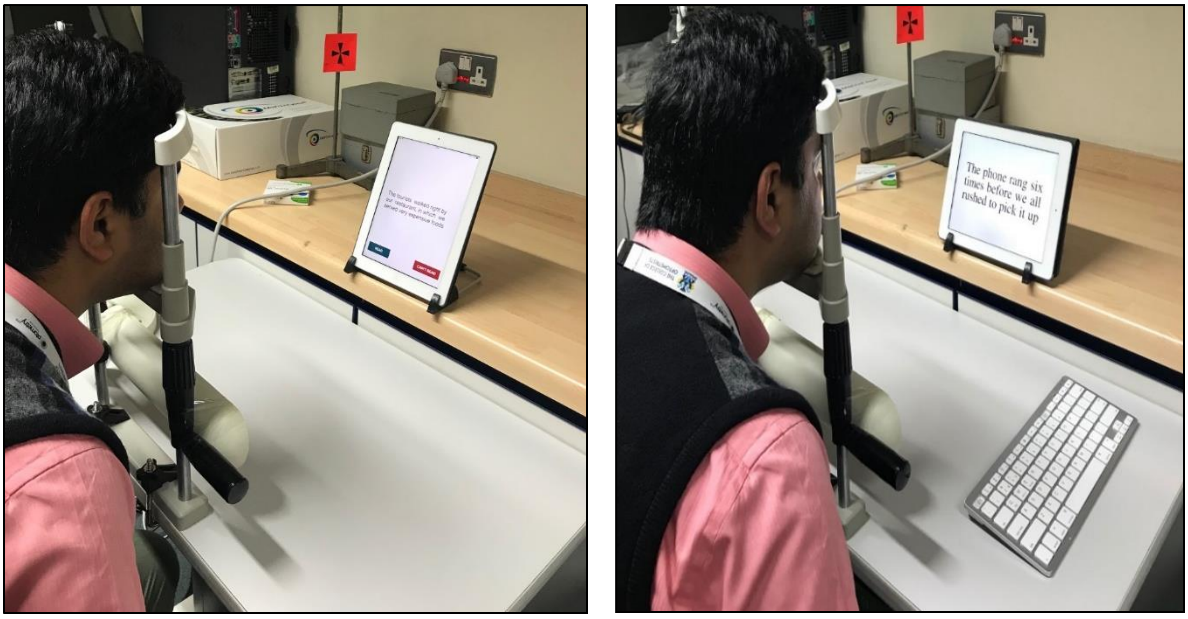


Figure 7.1: Experiments' setting procedure of the Radner reading test app (right) and MNread reading test app (left) (Baashen, 2020).

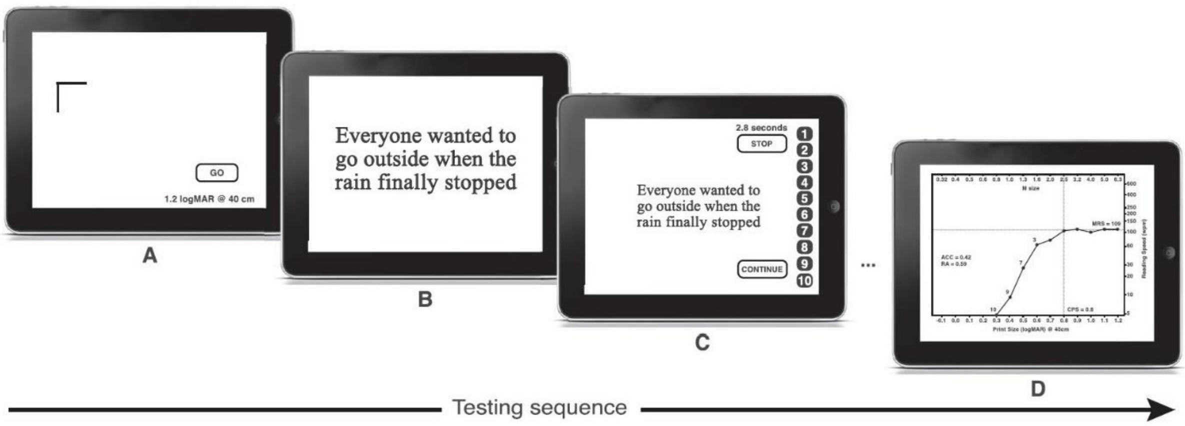


Figure 7.2: MNread app testing procedure. (A) preparation screen appeared before each sentence; (B) After clicking the 'Go' buttons by the examiner using the wireless keyboard, the sentence presented in the screen centre and the reading time was recorded simultaneously; (C) A score screen to enter the number of reading errors before launching the next sentence; (D) MNread data graphical plot appeared when the test was stopped. Image reproduced by (Calabrèse et al., 2018)

7.1.2 Radner app reading performance metrics estimation

In the Radner app, at the end of the test, the examiner can play back and listen to the sentences read by the subjects, tapping on the incorrectly read words to be accounted in the calculation of the reading acuity as the smallest print size that can be read + (0.005 X syllables of incorrectly read words). The Radner output was then simultaneously created for each participant and contained a table that represented the reading speed in wpm for each print size in addition to the graph which plotted the reading speed (wpm) against the print size (logRAD), see Figure 7.3. From this output, the examiner determined manually the CPS as the steepest decline slope and the MRS as the mean reading speed up to CPS. After tapping on the incorrectly read words, the number of the reading errors do not appear in the final output, but are used to calculate the reading acuity automatically.

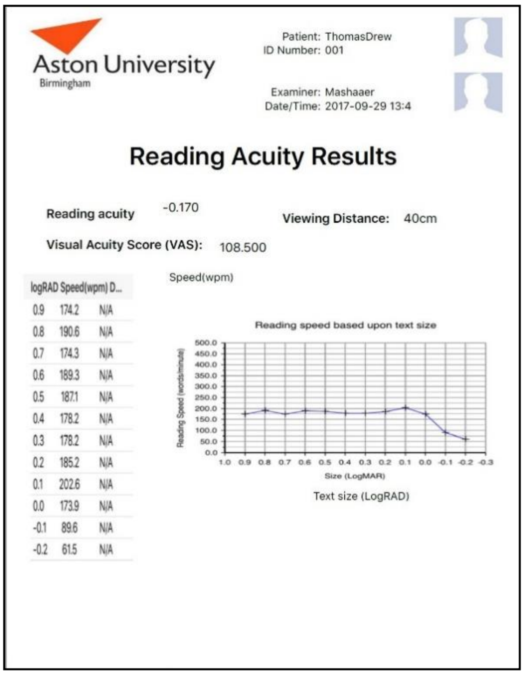


Figure 7.3: Radner reading speed test app output (Baashen, 2020).

7.1.3 MNread app reading performance metrics estimation

The MNread app has two modes of parameter estimation, automatic and manual. However, in this study, automatic estimation mode has been chosen for all the subjects; therefore, all the reading performance metrics were analysed by the app software. Once the test was complete, the MNread app created two files of output for each subject. The first file was the image of the graphical plot of reading speed in wpm against the print size (Fig.7.4), while the second was an Excel spreadsheet and contained the results of reading acuity, taking into account the number of reading errors, MRS, CPS and total number of reading errors. In addition, a detailed information about the reading time for each sentence was presented (Fig.7.5). The MNread app automatically measures the CPS as the print size at which the smaller print size is read at 1.96 SD slower than the mean of the preceding larger print sizes. The MRS is calculated as the mean reading speed of the sentences corresponding to CPS (Calabrese et al., 2014).

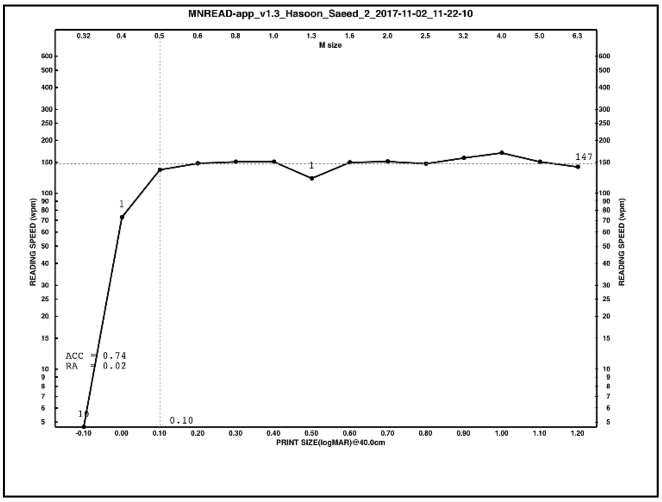


Figure 7.4: The MNread app curve output.

A	B	C
TimeStamp	2017-11-14_15-10-11	
Subject	Ayman 2	
ViewingDistance (cm)	40	
TestId	5	
NumberOfSentencesTested	14	
TotalReadingErrors	**0**	
ReadingAcuity (logMAR)	**-0.1**	
CriticalPrintSize (logMAR)	**0.1**	
MaxReadingSpeed (wpm)	**189.62**	
ReadingAccessibilityIndex	0.89495	
FittingMethod	Auto	
Polarity	Normal	
Font	TimesLT-Roman	
Software Version	1.3	
Device Name	Alec Kingsnorthâ€™s iPad	
Gamma	1.8	
Brightness	0.86	
Device Type	iPad3	4
iOS version	10.3.3	
Chart Id	English-3	
Notes	(null)	

SentenceId	PrintSizeAt40cm(logMAR)	PrintSizeAt40cm(M)	CorrectedPrintSize(logMAR)	FontSize(points)	x-height(points)	ReadingTime(seconds)	RawReadingSpeed(wpm)	ErrorCount	CorrectedReadingSpeed(wpm)	FullSentence
301-en	1.2	6.34	1.2	58.086	26.139	3.97	151.3	0	151.3	The phone rang six times before we all rushed to pick
302-en	1.1	5.04	1.1	46.14	20.763	3.07	195.5	0	195.5	She was waiting for over an hour before the doctor re
303-en	1	4	1	36.65	16.492	3.02	198.6	0	198.6	Instead of watching television we could play in your g
304-en	0.9	3.18	0.9	29.112	13.1	3.27	183.6	0	183.6	The boy carried the toy car in one hand for most of th
305-en	0.8	2.52	0.8	23.125	10.406	3.26	184.3	0	184.3	All of us ate dinner and then went over to see the ne
306-en	0.7	2	0.7	18.368	8.266	3.49	172	0	172	The children found a tiny chocolate pig inside each pa
307-en	0.6	1.59	0.6	14.591	6.566	3.5	171.6	0	171.6	The birds near your cabin made a lot of noise outside
308-en	0.5	1.26	0.5	11.59	5.215	3.02	198.6	0	198.6	We all thought that the grass needed to be cut twice
309-en	0.4	1	0.4	9.206	4.143	3.28	183	0	183	Never let your dogs play outside unless the cat has co
310-en	0.3	0.8	0.3	7.313	3.291	3.05	196.9	0	196.9	The tree in the front yard needs to be cut down by to
311-en	0.2	0.63	0.2	5.809	2.614	2.72	220.4	0	220.4	It was hard to know whether we were in the correct p
312-en	0.1	0.5	0.1	4.614	2.076	3.22	186.2	0	186.2	She could not sleep in the same room as the big scary
313-en	0	0.4	0	3.665	1.649	3.69	162.7	0	162.7	Have a nice time at the fair and be sure to come hom
314-en	-0.1	0.32	-0.1	2.911	1.31	4.91	122.3	0	122.3	There is a big river to cross before you reach the mou

Figure 7.5: The MNread app Excel spread sheet output.

7.1.4 Statistical analysis

The data were not normally distributed (Shapiro-Wilks test <0.05). Hence, the Spearman's correlation coefficient (r_s) was used to test the correlation between the two testing sessions for the reading performance metrics (RA, MRS and CPS) in the Radner and MNread tablet-based apps. Wilcoxon rank-signed test was conducted to determine if the repeated measurements differed significantly; the cut-off level for statistical significance was p <0.05. In addition, the Bland-Altman analysis was used to determine the agreement between the two sessions for each reading metric in the two apps. Similar statistical tests were used to compare the reading outcomes between the two iPad apps and to compare the tabled-based app with paper-based charts. Reliability of the reading apps was determined by calculating the Cronbach's alpha coefficient.

7.2 Results

7.2.1 Test-retest and inter-chart reliability of the Radner tablet-based app

As shown in Table 7.1, the repeated measurements of the reading acuity using Radner app resulted in a good correlation (r_s = 0.767) and agreement in the young group of subjects. In presbyopic subjects, although the Spearman's correlation was good (r_s = 0.727) between the repeated RA, the Bland-Altman agreement was low with 95% confidence interval CI: (-0.03; 0.10) logRAD. The repeated RA measurements showed mild correlation (r_s = 0.590) in cataract subjects and the agreement was low (Fig. 7.6a).

For the repeated MRS measurements, the Radner app resulted in good correlation and agreement in young and presbyopic subjects. However, the correlation for prespyopic subjects (r_s = 0.870) was higher than that in young subjects (r_s = 0.722). From Figure 7.6b, it can be seen the low Bland-Altman agreement in cataract subjects with 95% confidence interval CI: (-11.89; 2.35) wpm. Also, the correlation for measuring MRS was weaker in cataract subjects (r_s = 0.669) than normally sighted subjects.

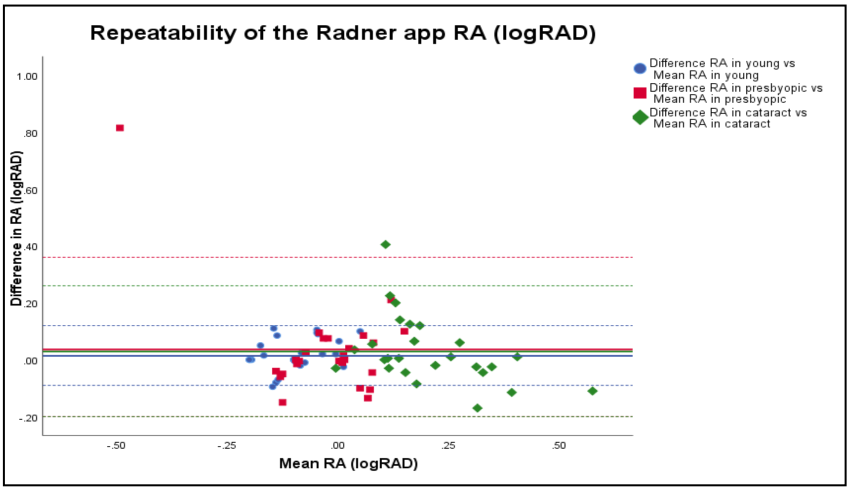
For the repeated CPS measurements, the Radner app revealed good repeatability results in the young group of subjects. Although, the CPS results was moderately correlated (r_s = 0.540) in presbyopic subjects, the Bland-Altman agreement was high (Fig. 7.6c). Very weak correlation and poor agreement were found between the two sessions for CPS results in cataract subjects.

In the young group of subjects, the reliability of the Radner app yielded a good Cronbach's alpha for RA (α = 0.891), MRS (α = 0.881) and CPS (α = 0.799). Excellent reliability analysis was found for MRS results (α = 0.930) in presbyopic subjects, but not for RA (α = 0.637) and CPS (α = 0.609). In the cataract group of subjects, the reliability of the Radner app was very poor for measuring CPS (α = 0.448) while it was higher for RA (α = 0.820) and MRS (α = 0.809).

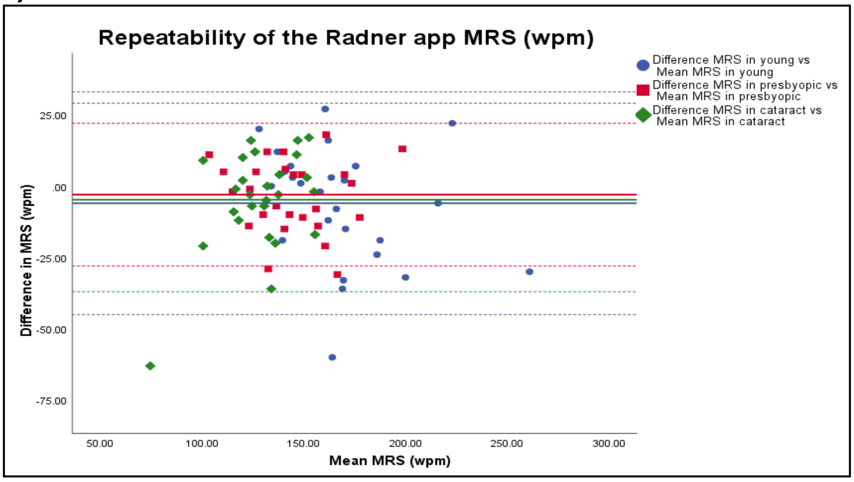
Table 7.1: Test-retest repeatability between the two testing sessions for measuring reading performance metrics using Radner app in three groups of subjects. (r_S) Spearman's correlation coefficient, *p*-value of Wilcoxon signed-rank test and Bland-Altman agreement between first and second visits.

	RA (logRAD)			MRS (wpm)			CPS (logMAR)		
Subjects	r_s	*p*-value	Agreement	r_s	*p*-value	Agreement	r_s	*p*-value	Agreement
Young	0.767	0.218	High	0.722	0.179	High	0.711	0.210	High
Presbyopic	0.727	0.428	Low	0.870	0.289	High	0.540	0.402	High
Cataract	0.590	0.390	Low	0.669	0.288	Low	0.273	0.193	Low

a)



b)



c)

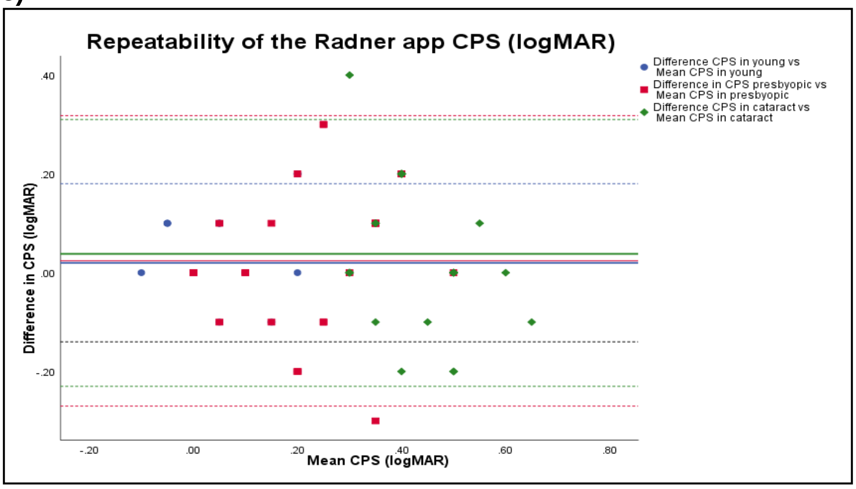


Figure 7.6: : Bland-Altman agreements in reading performance metrics, which are (a) reading acuity, (b) maximum reading speed and (c) critical print size between two testing sessions in Radner app in three groups of subjects: young n=29, presbyopic n=29 and cataract n=26. Agreements were low for all metrics in cataract subjects while high for all metrics in young subjects. There was high agreement for MRS and CPS in presbyopic subjects, but not for RA.

7.2.2 Test-retest and inter-chart reliability of the MNread tablet-based app

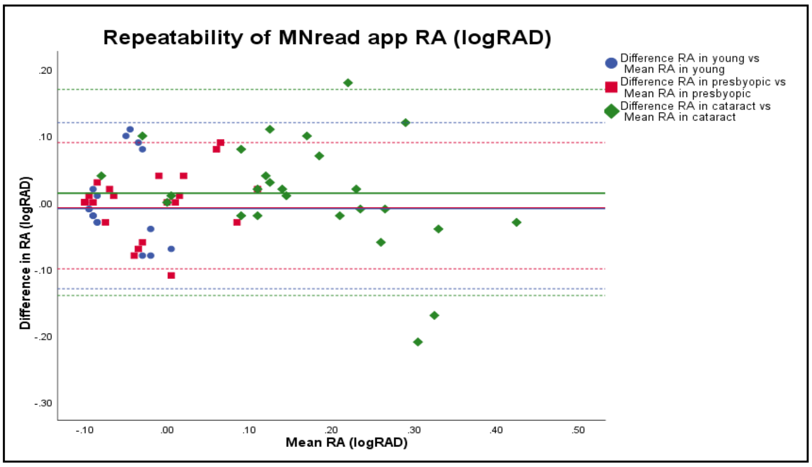
The results of the repeatability analysis for the MNread app are shown in Table 7.2. The repeated measurements of the reading acuity showed very poor correlation ($r_s = 0.267$) and low Bland-Altman agreement with 95% confidence interval CI: (-0.034; 0.015) logRAD in the young group of subjects. However, the RA repeatability results were good in presbyopic group of subjects. In cataract subjects, although high correlation was found for RA between the two sessions ($r_s = 0.851$), the agreement was low with 95% confidence interval CI: (-0.019; 0.047) logRAD (Fig. 7.7a).

MNread app showed good repeatability results (Fig.7.7b) for measuring the MRS in all groups of subjects. The highest correlation was in the presbyopic group ($r_s = 0.819$) followed by cataract group of subjects ($r_s = 0.736$) and younger subjects ($r_s = 0.706$). The MNread app results in very poor repeatability results for measuring CPS. The correlation between the two visits was ($r_s = 0.548$) in young, ($r_s = 0.262$) in presbyopic and ($r_s = 0.352$) in cataract subjects. Statistically significant differences $p < 0.05$ were found in the young and cataract group of subjects. Figure 7.7c shows the Bland-Altman agreements for measuring CPS using MNread app. The agreement analysis appeared to be unaffected and resulted in a good agreement of the repeated CPS measurements in young and presbyopic subjects, but not for cataract subjects. Turning now to the reliability results of the MNread app. In the young group of subjects, the Cronbach's alpha was unacceptable for measuring RA ($\alpha = 0.289$) and poor for measuring CPS ($\alpha = 0.551$). However, for MRS results, the reliability was good ($\alpha = 0.884$). In presbyopic and cataract subjects, the reliability were good and above 0.8 for measuring RA and MRS. For measuring CPS reliability, the result was unacceptable in the presbyopic group ($\alpha = 0.121$) and poor in cataract subjects ($\alpha = 0.528$). The number of the reading errors in the MNread app did not differ significantly between the first and second testing sessions in all groups of subjects $p > 0.05$.

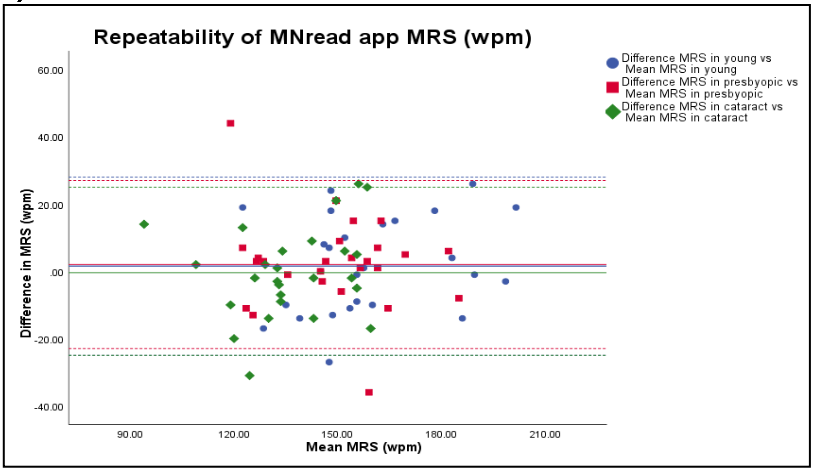
Table 7.2: Test-retest repeatability between the two testing sessions for measuring reading performance metrics using MNread app in three groups of subjects. (r_s) Spearman's correlation coefficient, *p*-value of Wilcoxon signed-rank test and Bland-Altman agreement between first and second visits.

	RA (logRAD)			**MRS (wpm)**			**CPS (logMAR)**		
Subjects	r_s	*p*-value	Agreement	r_s	*p*-value	Agreement	r_s	*p*-value	Agreement
Young	0.267	0.359	Low	0.706	0.569	High	0.548	0.386	High
Presbyopic	0.880	0.628	High	0.819	0.206	High	0.262	0.041	High
Cataract	0.851	0.220	Low	0.736	0.770	High	0.352	0.031	Low

a)



b)



C)

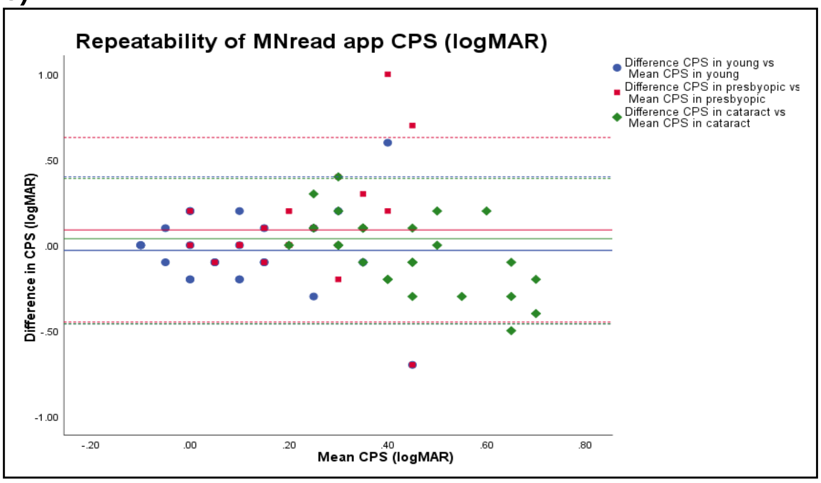


Figure 7.7: Bland-Altman agreements in reading performance metrics, which are (a) reading acuity, (b) maximum reading speed and (c) critical print size between two testing sessions in MNread app in three groups of subjects: young n=29, presbyopic n=29 and cataract n=26. Agreements were high for all metrics in young and presbyopic subjects, but not for RA in young subjects. There were low agreements for RA and CPS in cataract subjects, but not for MRS.

7.2.3 Comparison between Radner app and MNread app

The comparison between the two iPad app charts was used to determine the interest difference in the reading performance outcomes. In the young group of subjects, the RA (r_s 0.422) and the MRS ($r_s =$ 0.657) results showed weak to moderate correlation between the two apps. However, for measuring CPS, the Spearman's correlation was good ($r_s =$ 0.760). Wilcoxon rank-signed test found no statistically significant differences in reading metrics between the two apps in the young participants. Figure 7.8 shows the average reading speed for each print size using Radner app and MNread app.

In prebyopic participants, statistically significant differences were found between the two iPad apps for measuring the RA (median difference = -0.06 logRAD, p = 0.015) and MRS (median difference = 10 wpm, p = 0.008). As shown in Figure 7.9, the remarkable difference in the reading speed occurs in the very large (0.9 logMAR) and very small (0.0 logMAR) print sizes. The CPS measurement showed poor correlation ($r_s =$ 0.394) between the two apps.

In the cataract group of subjects, the results were similar to that found in presbyopic subjects. Statistically significant differences were found for measuring RA (median difference = 0.227, p = 0.002) and MRS (median difference = 10 wpm, p = 0.004). Also, the correlation between the two apps was poor for measuring CPS. In the cataract subjects, the MNread app resulted in faster reading speed compared to the Radner app (Fig. 7.10). Overall, these results indicate that different reading outcomes are revealed when using different tablet-based apps.

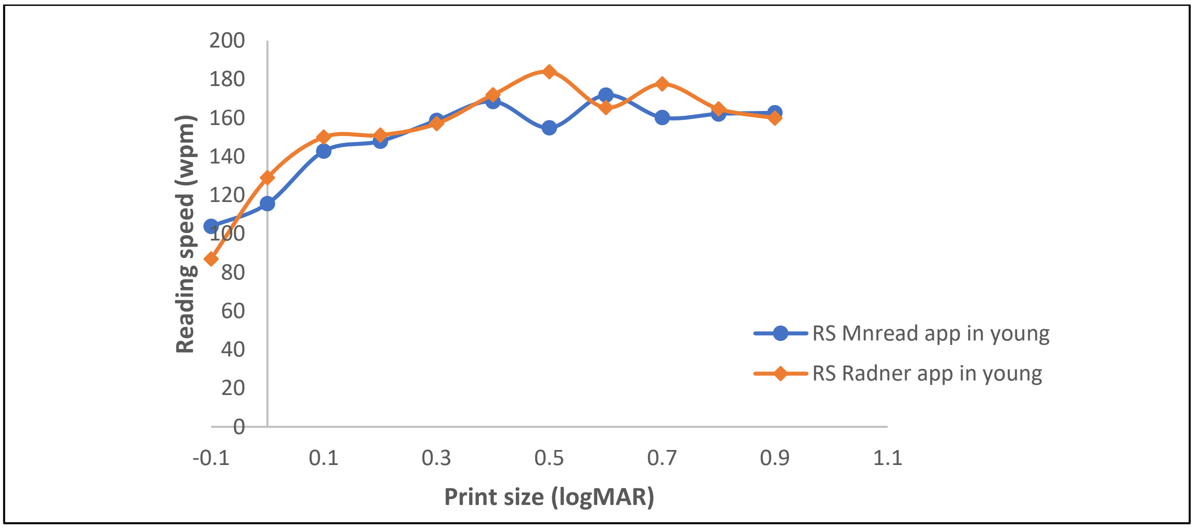


Figure 7.8: Mean reading speed at each print size for the two tablet-based reading apps in young subjects.

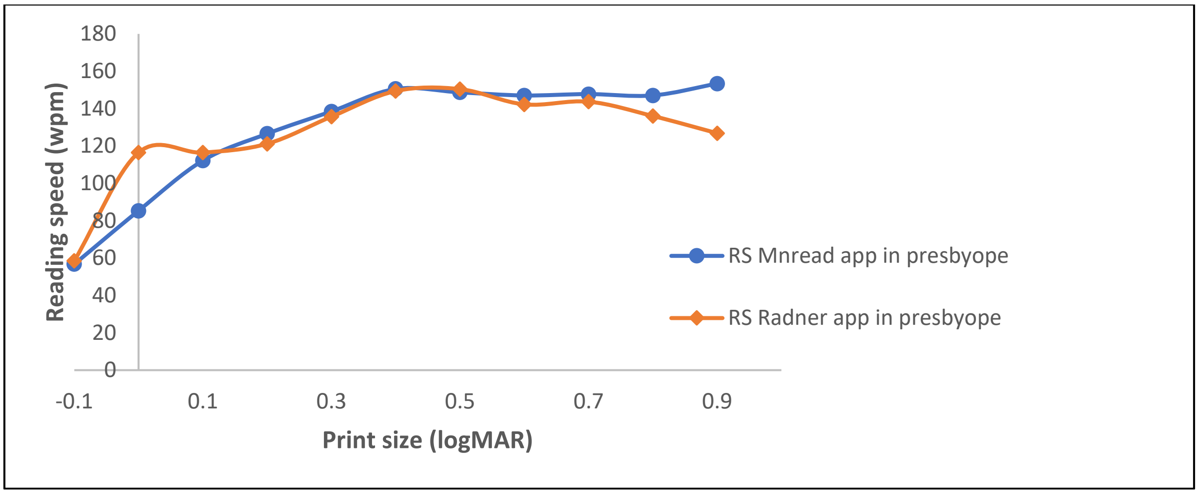


Figure 7.9: Mean reading speed at each print size for the two tablet-based reading apps in presbyopic subjects.

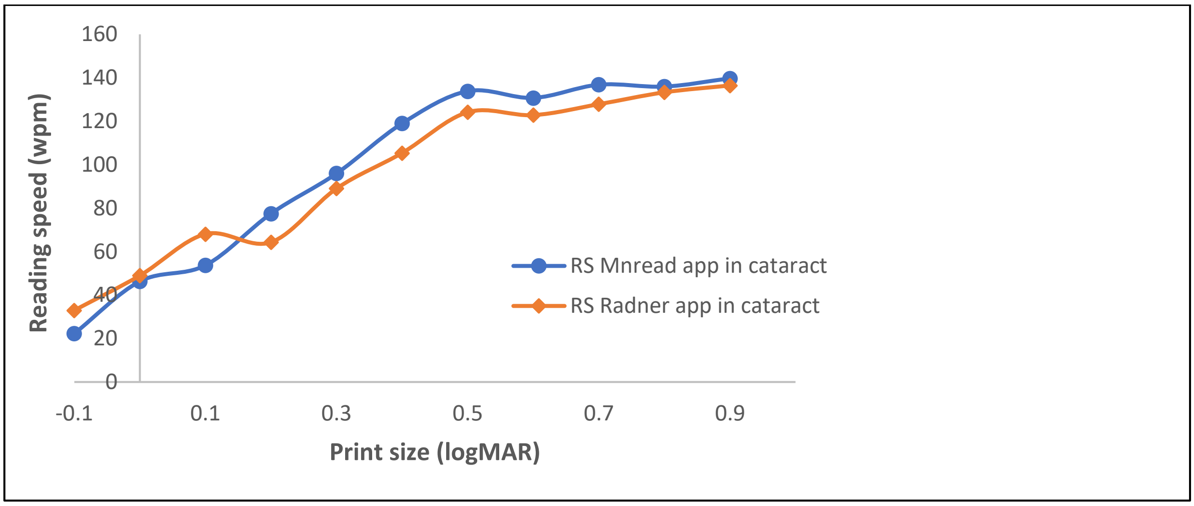


Figure 7.10: Mean reading speed at each print size for the two tablet-based reading apps in cataract subjects.

7.2.4 Comparing the reading performance on the MNread iPad app with the MNread paper chart

To test the validity and accuracy of the iPad apps, they were compared to the paper charts in the all group subjects. The data of the paper-based charts were normally distributed (Shapiro-Wilks test p >0.05), whilst, the data of the tablet-based app were not normally distributed (Shapiro-Wilks test p <0.05). Therefore, non-parametric analysis was used for the comparison.

The results of comparing the differences in the reading performance between the MNread app and MNread paper chart are presented in Table 7.3. The reading acuity results were moderately correlated in all groups of subjects. The differences between the app and the paper were statistically significant (p = 0.018), but small in the young and presbyopic subjects. These differences were only two letters in reading acuity. Compared to the MNread paper chart, the iPad app showed approximately similar RA results in the cataract subjects with high Bland-Altman agreement (Fig. 7.11a)

The MRS was significantly faster in paper chart (Fig.7.12) and weakly agreed with iPad app. The differences in MRS were (median difference = -46 wpm, 95% CI [-67, -36]) in young subjects, (median difference = -40 wpm, 95% CI [-46, -29]) in presbyopic subjects and (median difference = -30 wpm, 95% CI [-38, -24]) in cataract subjects. See Figure 7.11b for the MRS Bland-Altman agreement between the two types of MNread charts.

The CPS measurements showed poor Spearman's correlation coefficients in all subjects, ranging from 0.494 to 0.289. Statistically significant differences were found in presbyopic subjects and this difference was big and equal to one whole line (median difference = 0.1 logMAR, 95% CI [0.065, 0.244], p <0.001). The MNread app resulted in a larger CPS compared to the paper chart. The Bland-Altman plots of the CPS comparison between the MNread app and MNread paper chart are shown in Figure 7.11c.

Table 7.3: Difference between MNread app and MNread paper chart for measuring reading performance metrics in three groups of subjects. (r_s) Spearman's correlation coefficient, p-value of Wilcoxon signed-rank test and Bland-Altman agreement between MNread iPad app and paper chart.

	RA (logRAD)			**MRS (wpm)**			**CPS (logMAR)**		
Subjects	r_s	p-value	Agreement	r_s	p-value	Agreement	r_s	p-value	Agreement
Young	0.662	0.018	Low	0.457	0.000	Low	0.494	0.335	High
Presbyopic	0.733	0.018	High	0.688	0.000	Low	0.310	0.000	Low
Cataract	0.683	0.976	High	0.728	0.000	Low	0.289	0.217	High

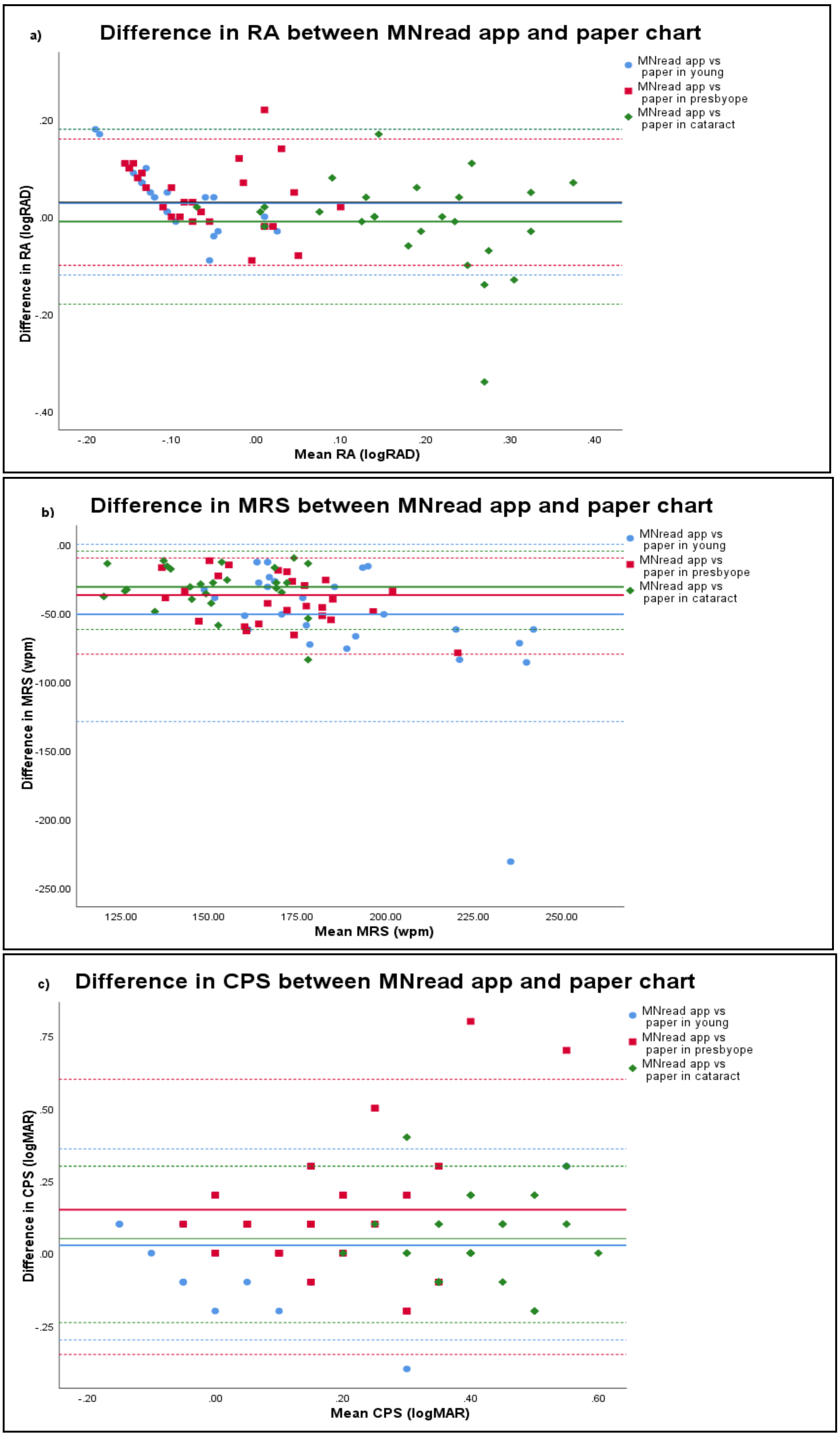


Figure 7.11: Bland-Altman agreement of difference in reading performance metrics against the mean comparing MNread app with MNread paper chart in three groups of subjects. (a) For RA, the agreements were high in presbyopic and cataract subjects, but low in young subjects. (b) The MRS agreements were low for all the subjects. (c) High agreement for CPS was found in young and cataract subjects, but not for presbyopic subjects.

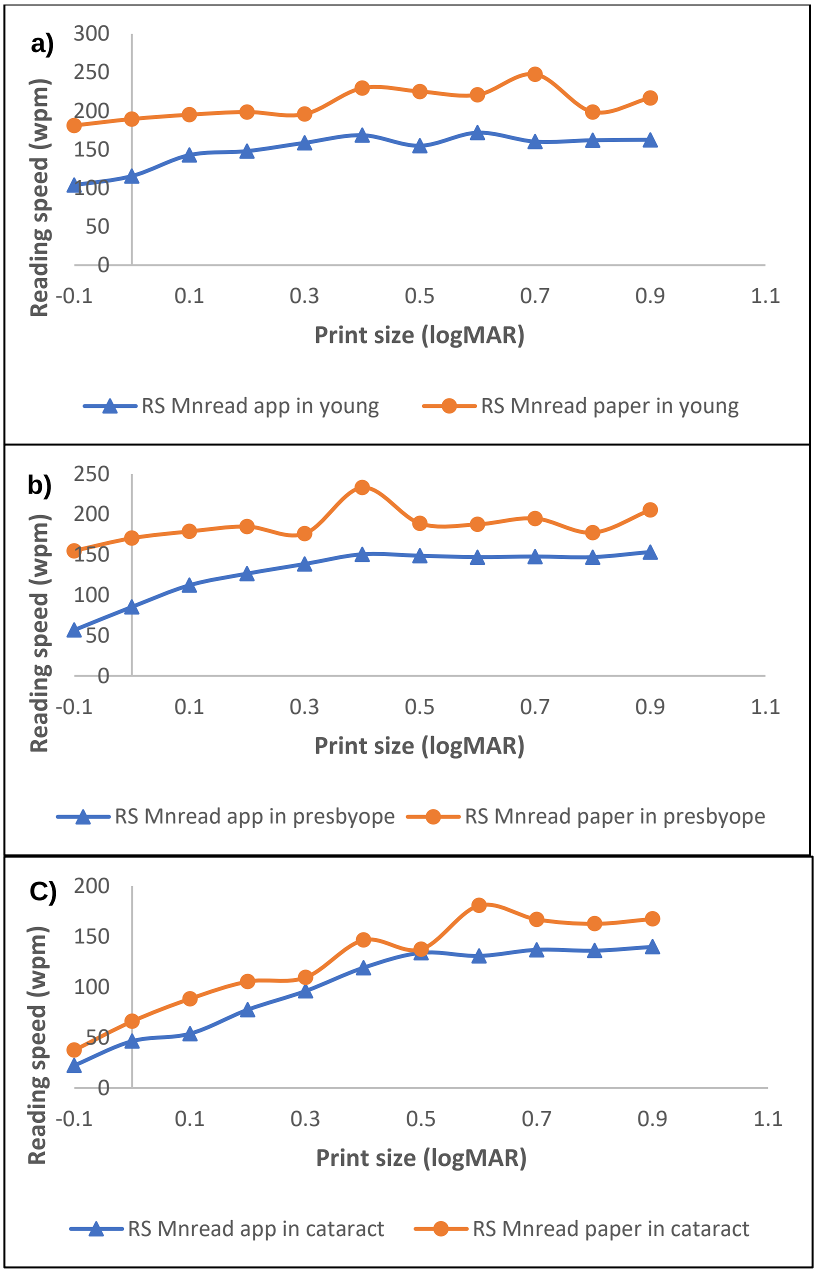


Figure 7.12: Mean reading speed at each print size for MNread app and MNread paper charts in (a) young subjects, (b) presbyopic subjects and (c) cataract subjects. Reading speed in the paper chart was faster than the app.

7.2.5 Comparing the reading performance on the Radner iPad app with the Radner paper chart

Table 7.4: Difference between Radner app and Radner paper chart for measuring reading performance metrics in three groups of subjects. (rs) Spearman's correlation coefficient, p-value of Wilcoxon signed-rank test and Bland-Altman agreement between Radner iPad app and paper chart.

	RA (logRAD)			MRS (wpm)			CPS (logMAR)		
Subjects	r_s	*p*-value	Agreement	r_s	*p*-value	Agreement	r_s	*p*-value	Agreement
Young	0.674	0.002	High	0.560	0.000	Low	0.682	0.850	High
Presbyopic	0.749	0.436	High	0.741	0.000	High	0.415	0.545	High
Cataract	0.699	0.116	High	0.716	0.000	High	0.420	0.811	High

Table 7.4 above presents the summary statistics for comparing the reading performance metrics between Radner iPad app and Radner paper chart in young, presbyopic and cataract subjects. For measuring the reading acuity, high Bland-Altman agreement between the Radner app and paper was found in presbyopic and cataract subjects, with no statistically significant difference. Also, the Spearman's correlation for RA was relatively high ($r_s =$ 0.749) in prespyopic and ($r_s =$ 0.699) in cataract subjects. However, in the young group of subjects, the difference was statistically significant (median difference = 0.095 logRAD, 95% CI [-0.08, -0.01], p = 0.002). Although, this difference was significant, the Bland-Altman plot showed a good agreement (Fig. 7.13a).

For measuring MRS, the Radner paper chart always resulted in a faster reading speed compared to the iPad app in all groups of subjects p <0.05. The discrepancies of the reading speed between the two types of Radner charts are shown in Figure 7.14. The differences in MRS were (median difference = -22 wpm, 95% CI [-44, -21]) in young subjects, (median difference = -37 wpm, 95% CI [-45, -30]) in presbyopic subjects and (median difference = -28 wpm, 95% CI [-39, -22]) in cataract subjects. However, the correlations and agreement were high in presbyopic and cataract subjects, but not for the young subjects (Fig. 7.13b).

The Radner app seems to match and correlate well with the paper chart for measuring the CPS. Although, the Spearman's correlations were low in presbyopic and cataract subjects, r_s = 0.415 and 0.420, respectively, the medians of the CPS were exactly the same in the Radner app and Radner paper chart. The Bland-Altman analysis resulted in high agreement in all groups of subjects without any statistically significant differences (Fig. 7.13c).

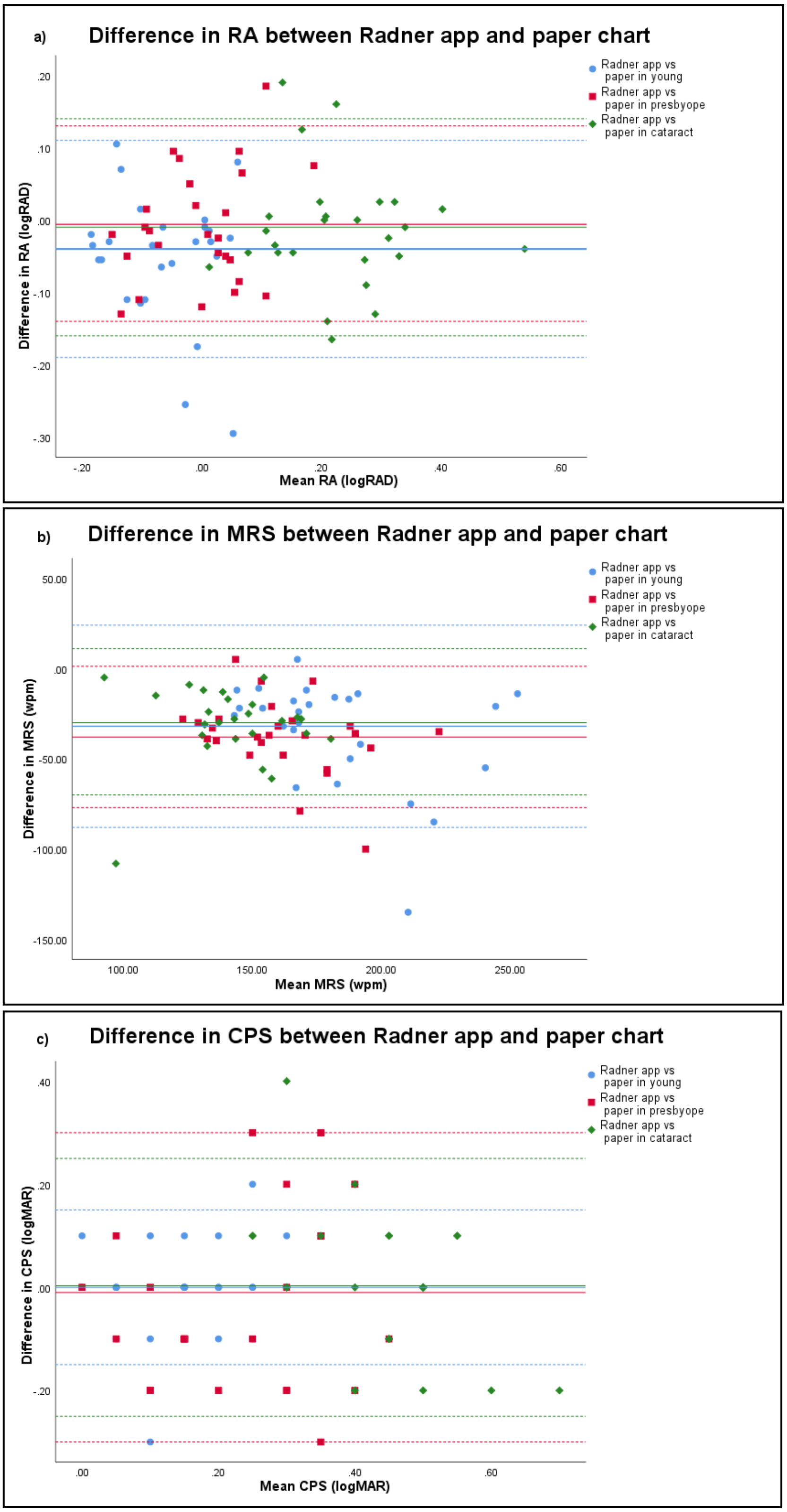


Figure 7.13: Bland-Altman agreement of difference in reading performance metrics against the mean comparing Radner app with Radner paper chart in three groups of subjects. (a) For RA, the agreements were high in all subjects. (b) The MRS agreements were low in young subjects, but high in presbyopic and cataract subjects. (c) High agreement for CPS was found in the three groups.

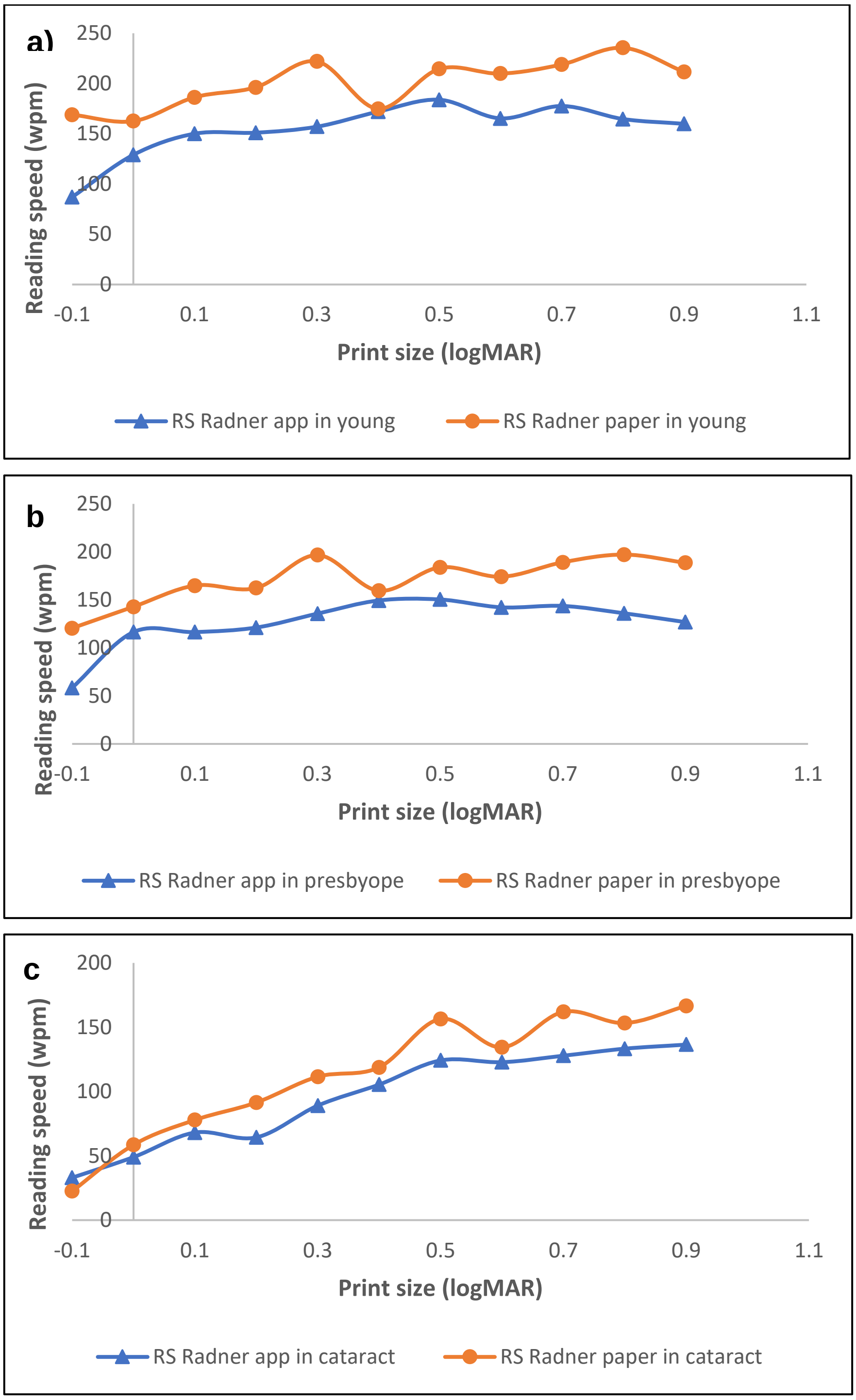


Figure 7.14: Mean reading speed at each print size for Radner app and Radner paper charts in (a) young subjects, (b) presbyopic subjects and (c) cataract subjects. Reading speed in the paper chart was faster than the app.

7.3 Discussion

7.3.1 Test-retest and inter-chart reliability of the Radner and MNread tablet-based apps

Using the tablet-based reading test in clinical practice has many advantages, including: uniform background illumination, the quality of the test cannot be degraded with time as in a traditional paper chart, automated timing and voice detection, automated reading speed graph plotting, standardising the methods for scoring the reading performance metrics, saves more time and is portable.

In previous chapters, the repeatability and reliability of the Radner and MNread paper-based charts have been investigated. The tablet-based apps contain sentences that are selected based on the traditional paper charts. The results of this study showed that, in the young group of subjects, the Radner app resulted in high agreement between the two testing sessions for measuring RA, MRS and CPS. Also, the reliability analysis resulted in a good Cronbach's alpha range from 0.799 to 0.891. With respect to the repeatability results of the Radner paper chart in young subjects that have been evaluated in Chapter 3, it can be suggested that the Radner app has better repeatability results than the paper chart. With regard to the characteristics of the MNread app in the young group, the repeatability and the reliability were very poor for measuring the RA and CPS. This contrasts with that found in the MNread paper chart, which resulted in a better repeatability and agreements between the two testing sessions.

In the presbyopic subjects, although the correlation (r_s = 0.727) between the two testing sessions was good with no statistically significant difference for measuring RA using the Radner app, the Bland-Altman agreement was low and the reliability was questionable (α = 0.637). The possible explanation for this may be related to the self-administration of the Radner app test. Even though the participants were correctly instructed about how to run the test, some ended the test early because they felt this was the smallest print size they could read without making an attempt to read the next print size sentence. In such cases, it can be thus suggested that the reading test app is better administered and controlled by the examiner to avoid any variability in the response action or response time from the participants. In presbyopic subjects, the Radner app has better repeatability results than that found with the paper chart (Chapter 4). On the other hand, the MNread paper chart resulted in better repeatability results than the MNread app. In the MNread app, the CPS measurements were poorly correlated between the two testing sessions (r_s = 0.262).

In the cataract group of subjects, the tablet-based apps revealed low agreements and bad repeatability results for measuring all the reading performance metrics, except for the MRS repeatability in the MNread app which was good. The most likely cause of these results may relate to the maximum brightness of the iPad screen, that may produce a glare effect in cataract subjects. Therefore, the use of antiglare screen protector is recommended to avoid any reflection on the screen and help to improve the reading performance. It should be noted that characteristics of the MNread app in terms of repeatability and reliability were strong for measuring the MRS only in all groups of subjects. This observation may support the future research studies that are concerned about measuring the reading speed only using the MNread app, which provides a more automated uniform method for calculating MRS.

To conclude this section, using the tablet-based apps to measure the reading performance metrics in the cataract patients should be interpreted with caution. The Radner app is a valid and repeatable test of assessing the reading performance in normally sighted individuals.

7.3.2 Comparison between Radner app and MNread app

The Radner app contains two versions with repeated sentences within the chart, while the MNread app contains five versions without any sentence repetition and this gives the MNread app an advantage to overcome the learning effect and avoid memorisation.

The results showed that comparing the two tablet-based apps resulted in different reading performance outcomes. The CPS measurements were weakly correlated between the two apps in the young and presbyopic subjects, the most likely cause being that the MNread app provided the automated calculation of the CPS, whereas it was calculated manually in the Radner app. In normally sighted individuals, using the graphical plot of the reading speed as a function of the print size may be sufficient to extract the CPS visually by the examiner. However, this can be difficult to extract for individuals with visual impairment due to the variability of the curve information (Baskaran et al., 2019). In such cases, using the automated rater to calculate CPS, as in the MNread app, may reduce the variability, avoid subjective decision and unify the scoring rule method.

In addition, the results of the reading acuity and MRS showed statistically significant differences between the Radner app and MNread app. As it has been proven in the previous chapters (3 and 4), different reading performance outcomes are revealed when using different paper reading charts; similarly, the use of different tablet-based reading tests resulted in different reading outcomes. This is an important issue to consider in clinical practice or research studies that use the reading tests to assess the outcome of the treatment for the eye disease or to evaluate the efficacy of the presbyopia correction.

7.3.3 Comparing the reading performance on the tablet-based apps with the paper-based charts

The paper-based charts measured a markedly faster reading speed than the equivalent tablet-based apps, presumably because of the latency to stop the timer in the apps (timing method). In the paper charts, audio recording was used, which started when reading the first sentence in the chart and stopped after reading the last sentence in the whole chart. The examiner then played the recording and broke down the reading time for each sentence. In the Radner app, the young and presbyopic participants were administered the test and pressed the 'Read' button after each sentence to record the reading time; thus, the slower reading speed in the Radner app may be due to the latency in the response time. Although, the Radner app reading test was administered by the examiner in the cataract group of subjects, the results were the same and the Radner app resulted in significantly slower reading speed. However, this outcome is contrary to that of Kingsnorth and Wolffsohn (2015) who found that the reading speed of the young subjects was faster in the Radner app compared to the paper chart. This contradictory result of their study may be due to performing the Radner app test twice in the same period, that could enhance the practice effect and this may lead to a faster reading speed, especially given that the Radner app has a sentence repetition and all the participants were young.

The Radner app and Radner paper chart showed a good matching for measuring CPS and resulted in high Bland-Altman agreement with no statistically significant differences in all groups of subjects. This high agreement can be explained by using the same scoring method for calculating CPS. As mentioned earlier in the methods section, the Radner app does not calculate the CPS automatically and rather displayed the curve of the reading speed as a function of print size and then the examiner has to visually estimate the CPS as the steepest decline slope. The exact scoring method was used for the paper charts. On the other hand, the correlation of CPS results between the MNread app and the MNread paper chart was very poor. This result is again likely to be related to the different CPS scoring method used in MNread app as it was calculated automatically as the print size at which the smaller print size was read, at 1.96 SD, was slower than the mean of the preceding larger print sizes. This finding has important implications for standardising the scoring rule method for calculating the reading performance metrics between the reading charts.

The results showed a small but significant difference in the reading acuity results between MNread app and MNread paper chart in young and presbyopic subjects. It seems possible that the different print size ranges contributed to this difference. In the MNread app, the print size ranges from 1.2 to -0.1 logMAR, while in the MNread paper chart the print size ranges from 1.3 to -0.5 logMAR. The near visual acuities measured with near ETDRS in young and presbyopic participants were (mean, -0.06± 0.08) logMAR and (-0.01± 0.08) logMAR, respectively, which means that they have ability to read further than the smallest print size in the MNread app. This also explained why there is no significant difference in reading acuity between MNread app and paper chart in cataract subjects, as their ability to read print size smaller than -0.1 logMAR was rare and their near visual acuity was (mean, 0.23± 0.14) logMAR. On the other hand, high agreement was found between the Radner app and Radner paper chart for measuring the reading acuity in all subjects and this is because the smallest print size in both charts is the same and equal to -0.2 logMAR. It has been recommended to change the viewing distance when using the MNread app to shorter or longer to compensate the paper chart print size range (Calabrèse et al., 2018). Longer viewing distance is recommended to use with normally sighted subjects to extend the range of the print size at the smaller end while shorter viewing distance is recommended to use for the individuals with ocular disease to get larger print sizes.

The result of comparing the tablet-based apps with the paper-based charts suggests that the Radner app matches and correlates with the paper version better than the MNread app.

7.4 Conclusion

The results of MNread and Radner tablet-based apps are not interchangeable with a paper-based chart. The comparison between the two apps found that, generally, the Radner reading speed test app is reliable and repeatable for measuring the reading performance metrics in normally sighted individuals. The question raised for using the reading test app for patients with ocular diseases such as cataract is whether this would be a fruitful area for further improvement of the digital eye test. The Radner app has several advantages over the traditional chart, including that it is quick to perform, has accurate timing measurement, and uniform automated methods for estimation the reading performance and this will help to standardise the reading assessment.

Chapter 8

Best methods for measuring the reading performance in clinical practice

8 Introduction

This chapter has been divided into four experiments concerned with what is the best methodology to assess reading performance accurately and effectively, either in clinical practice or in research studies, where measuring reading performance is critical.

8.1 The accuracy of reading speed measurement by stopwatch versus audio recording method

Important considerations regarding comparing the reliability of reading test charts between the studies are the patient characteristics (Radner et al., 2002) and measurement-related factors (Lovie-Kitchin and Brown, 2000), both of which may influence the variability of the reading performance results. The measurement-related factors that need to be described include viewing distance, illumination level and the recording method.

If reading speed is calculated by determining the time period, the person needs to read a specific sentence. Thus, the accuracy of the time measurement method is a critical element for analysis of the reading speed. Bailey and Lovie (1980) found that it is possible to calculate the reading acuity and the reading speed in one simultaneous examination using a stopwatch. Their method has also been applied to other reading charts, such as Radner and MNread. The majority of the studies that have evaluated reading performance used the stopwatch method for recording the reading time of each sentence in the reading chart in one simultaneous examination, whilst digital and audiotape recordings were used to measure the reading speed after the test in studies such as Radner et al. (2002), Maaijwee et al. (2008), Burggraaff et al. (2010a), Subramanian and Pardhan (2009), Maaijwee et al. (2008b), Radner et al. (2002). Also, the reading performance metrics calculation in the previous chapters of this thesis was performed after the test session using the audio recording method. Since the accuracy of the reading time calculation is crucial for measuring the reading performance metrics, this study compared the traditional stopwatch measurements with audio recording measurement of the reading speed.

8.2 Eye gaze position and reading speed

Reading posture may affect reading effectiveness as well as accommodation demand, convergence, retinal image quality and comfort during reading (Scheiman and Wick, 2008). During the reading performance test, it is worth noting how the patients actually look at the reading chart while reading. Most practitioners evaluate reading speed by one of two methods: either putting the chart on the reading stand in front of the patient's eyes at the recommended chart distances and letting the patient focus straight ahead and use their central position of gaze while reading; or when the patient holds the reading chart in an armchair setting and using the reading position with downgaze. In both settings, patients assume a variety of gaze angles and head postures. Some of the basic parameters, such as eye position of gaze during the reading, have no empirical support. Therefore, one aim of this chapter is to evaluate reading speed in these two settings to contribute knowledge about the standardized way for measuring the reading speed in routine clinical practice or in research studies.

8.3 Correlation of measures of vision and reading performance

Contrast sensitivity test provides useful information about visual function which may not be detected when using the high contrast visual acuity test (Hirvelä et al., 1995). However, measuring the full contrast sensitivity function is less common to perform in clinical practice than in research studies because it is a time-consuming procedure (Pelli and Bex, 2013). Therefore, in clinical setting, contrast sensitivity test is usually measured at one spatial frequency by a printed chart such as Pelli-Robson chart or near Mars Letter Contrast Sensitivity chart (Pelli and Robson, 1988). While printed contrast sensitivity charts are still in use, computer-based testing is becoming more popular and allows to generate a wide range of stimuli and maintain the good quality of the test over a long period of time. An example of this is the liquid crystal display monitor (LCD) which is commonly used in clinical practice and includes a computer-based contrast sensitivity test measuring at distance of 4m. The present study evaluated the agreement between the near contrast sensitivity measured with Mars test and distance contrast sensitivity measured with LCD system. In addition, the study determines the ability of both tests to discriminate between normally sighted subjects and the patients with cataract.

An understanding of how the visual examinations influence the reading performance is important and may provide useful information about how the visual functions, such as reading, are correlated to primary vision examination outcomes. Visual acuity is the most widely used test to evaluate visual function. However, some visual functions, such as night driving, mobility, reading and ability to discriminate between targets, cannot be detected by measuring visual acuity alone (Lott et al., 2001b). Several everyday activities depend on both visual acuity and contrast sensitivity. However, some of the visual tasks may be related more to visual acuity, such as tasks that need high resolution, or related more to contrast sensitivity, such as face recognition, mobility and night driving (Alexander et al., 1988). Leat and Woodhouse (1993) suggested that the reading performance with low vision aids is better predicted by measuring contrast sensitivity than visual acuity. However this finding is contradicted by other studies (Legge et al., 1992, Tejeria et al., 2002).

Reading is an important skill and practitioners need to estimate how well the patients perform the reading task. Therefore, it is useful to know some of the routine clinical data used in predicting the reading performance, especially when it is not practical to measure the reading performance directly due to the length of the series of eye examination tests. Therefore, one objective of this study was to identify which measures of vision correlate well with reading performance metrics measurements.

8.4 Repeatability of near visual acuity measurement with different optotypes

The gold standard for VA measurements is any chart that adheres to the Bailey-Lovie chart design principles irrespective of testing distance or the type of optotype (Bailey and Lovie, 2013). Bailey and Lovie design principles were developed to overcome the problems arising from Snellen charts and have certain advantages, such as the logarithmic progression from one line to the next (Bailey and Lovie, 1976), same number of letters on each line and the spaces between the letters and lines are proportional to the letter size. The ETDRS and Bailey-Lovie charts were widely used in research studies or in clinical practice. Since then, a number of new charts have been developed based on the Bailey and Lovie chart design with different optotypes, including Landolt C, tumbling E, letters, numbers and words. These different optotypes use a logMAR progression of 0.10 log units, five optotypes per line and proportional spacing with letter size. It has been addressed that the logMAR progression charts have many advantages over Snellen and other notations (Bailey and Lovie, 1976). Despite the similarities of the logMAR charts designs, it cannot be assumed that different optotypes will give identical VA scores. A literature review by Williams et al. (2008) found a lack of uniformity of visual acuity testing in published studies but they were rarely described in detail, which may make it difficult for other researchers to duplicate the study.

For distance VA, Landolt C optotype was adopted as standard of reference by the National Academy of Sciences-National Research Council (NAS-NRC). The advantages of the Landolt rings are international recognition, equal difficulty of each ring and minimum artefact produced by rotating the ring around a fixed point (Ruamviboonsuk et al., 2003). However, very little is known about the standard optotype preference for VA measurement at near. One purpose of this study is to assess the near visual acuity measured with different logMAR optotypes (Landolt C, tumbling E, letters and numbers) in terms of test-retest and inter-chart reliability. In addition, to compare the near visual acuity measured with different logMAR optotypes between each other and between the reading acuity measured with word reading chart (Bailey-Lovie word reading chart).

8.5 Methods

8.5.1 The accuracy of reading speed measurement by stopwatch versus audio recording method

Young normally sighted and presbyopic subjects were chosen to participate in this comparison with exclusion of cataract subjects because of the expected variability in the results. Therefore, the total number of participants was 58 subjects with age ranged from 18 to 70 years old.

All the participants read the 19 sentences in the MNread reading chart aloud, sentence by sentence. The audio recordings were made with computer and, at the same time, the examiner measured the reading time of each sentence manually with a digital stopwatch. The examiner started measuring the reading time at the vocal onset (premovement of the lips) as recommended by Radner et al. (2017). Reading errors were noted by marking the wrong word on the MNread scoring sheet in the same clinical setting. Afterwards, GoldWave Inc (St. John’s, Newfoundland, Canada) software was used for playing the audio recordings and measuring the reading time and the reading errors. This programme allows the examiner to play the recordings at slow speed and display the sound waves visually. Thus, this study contains two measurements per reader and 116 measurements for the whole study. All the measurements were performed by one examiner.

Statistical analysis compared the maximum reading speed (wpm), reading errors and reading time (seconds) obtained with the stopwatch with those of the audio recording method. The data of the reading time were normally distributed (Shapiro-Wilks >0.05) and paired sample t-test was used to analyse the significant difference between the two recording methods. As the MRS and the reading errors data were not normally distributed (Shapiro-Wilks <0.05), a non-parametric Wilcoxon signed-rank test was, therefore, used, and two-sided p value <0.05 was considered to be statistically significant. To measure the agreements of the reading time and MRS between the stopwatch and audio recording methods, Bland-Altman analysis was used.

8.5.2 Eye gaze position and reading speed

Twenty-nine young normally sighted and 26 presbyopic subjects participated in this experiment. Three presbyopic subjects were excluded because it was not possible to adjust the straight-ahead reading position on the reading desk due to their body height. The Colenbrander reading chart was used to measure the reading speed. The test was performed binocularly at 40cm. The subjects read the sentences under two different reading conditions:

1- Head downward position in armchair setting without table or desk, this condition will be referred to as "read in armchair"
2- Straight ahead position looking at the chart on the reading stand at the desk, this condition will be referred to as "read at disk"

In each reading condition, the subjects were asked to read the sentences aloud as fast as they could without correcting the reading mistakes. All testing procedures were monitored with audio recording to measure the reading time for each sentence and the reading errors. In read at disk condition, the reading distance of 40cm was controlled using a headrest for the forehead, while, in read in armchair condition, the reading distance was measured constantly during the test using a centimetre ruler (Fig. 8.1).

The data of the young group of subjects were not normally distributed (Shapiro-Wilks <0.05), therefore, Spearman correlation coefficient and Wilcoxon signed-rank test were used to analyse the statistically significant difference in the reading speed and reading errors between read in armchair and read at disk conditions. The data of the presbyopic group of subjects were normally distributed and parametric tests were used; a two-sided p value <0.05 was considered to be statistically significant.

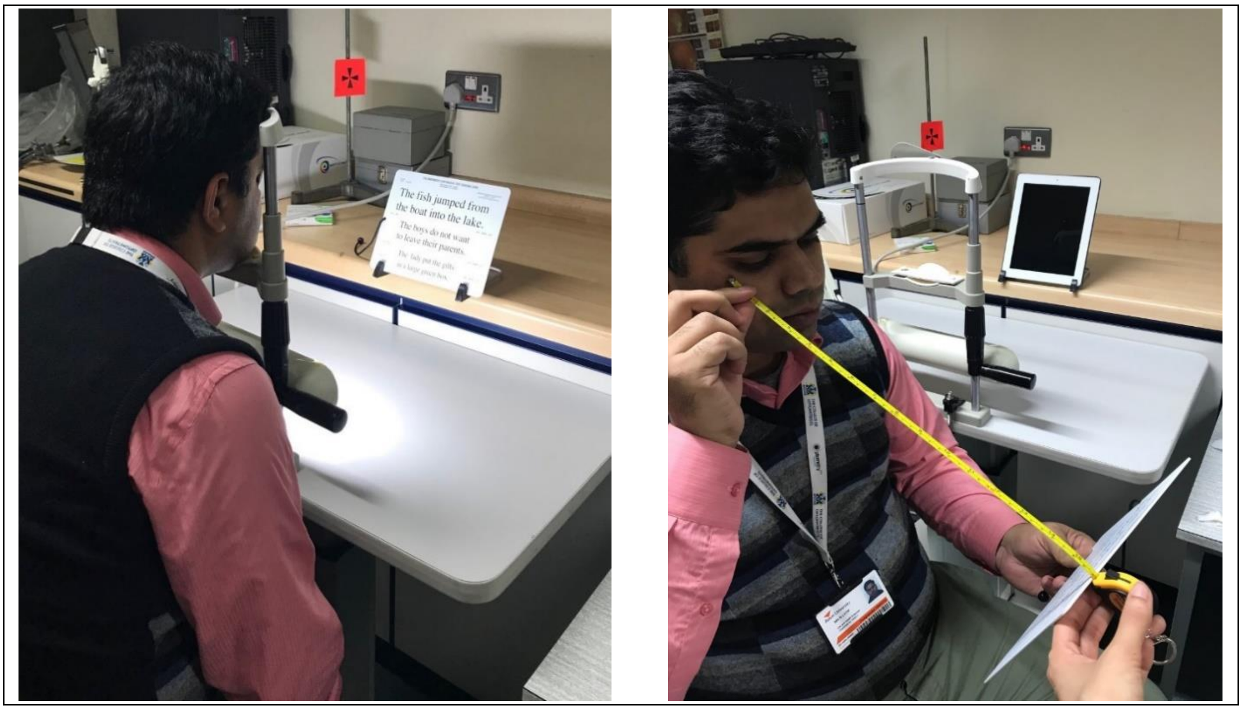


Figure 8.1: Measurement of the reading speed in two different reading conditions (right) read at desk condition (left) read in armchair condition

8.5.3 Correlation of measures of vision and reading performance

Three groups of subjects were recruited in this experiment (young, presbyopic and cataract subjects). Distance visual acuity (VA), near visual acuity, distance contrast sensitivity (CS) and near contrast sensitivity were measured in all subjects. Measurements were done monocularly in all groups of subjects. To determine the reading performance metrics, which were reading acuity, maximum reading speed and critical print size, the Bailey-Lovie unrelated word reading chart was used. This reading chart was chosen in this experiment because it revealed a good repeatability and reliability characteristics for measuring reading performance based on the results of the previous chapters (Chapters 3, 4 and 5).

Corrected distance visual acuity was measured using a standardised Early Treatment Diabetic Retinopathy Study (ETDRS) chart displayed by software CSO vision chart v 2.6.0 on an LCD screen at distance of 4m. Near visual acuity was measured using ETDRS near vision chart (Precision Vision) at 40cm with optimal near correction.

Distance contrast sensitivity was measured using the LCD system, CSO Vision Chart (Mod CVC02, software CSO Vision Chart v 2.6.0 CSO, Florence, Italy) with distance habitual correction. The LCD system produced a contrast sensitivity chart following the same criteria as the Pelli-Robson chart (Pelli and Robson, 1988). The LCD CS test contains 16 groups of three Sloan letters with the same contrast. The maximum contrast is 93% and the minimum contrast is 0.04%, decreased contrast factor is 0.15 log units. Unlike the Pelli-Robson chart, the triplets are presented one at a time in the LCD CS test due to the limited screen size. Distance CS test was performed at testing distance of 4m. Recently, it has been found that the LCD CS test has a good accuracy and repeatability, but is not interchangeable to the printed Pelli-Robson chart (Zeri et al., 2018).

Near contrast sensitivity was measured using Mars Letter Contrast Sensitivity test developed by Mars Perceptrix Corporation in 2004 (Arditi, 2005). The test was performed with optimal near correction, if needed. The Mars CS test is a portable chart used at the test distance of 50cm. It consists of 48 letters arranged in eight rows, each having six letters. Its maximum contrast is 91% and the minimum contrast is 1.2%, the contrast decreased by factor of 0.04 log units. The Mars Contrast Sensitivity test showed a good agreement and similar test-retest repeatability with the Pelli-Robson test. Several advantages result from the small size of the Mars CS test which make it suitable for near testing and portable use (Dougherty et al., 2005).

For both contrast sensitivity tests, subjects were instructed to read the letters and encourage to guess and observe the letters for at least 20 seconds, to allow the time for perception at threshold (Elliott et al., 1991). Letter-by-letter scoring method was used for both near and distance CS tests where the values of 0.05 log units and 0.04 log units were given per correct letter for LCD CS and Mars tests, respectively.

The data in this experiment were not normally distributed (Shapiro-Wilks <0.05). The strength of the relationship between distance CS and near CS was evaluated using Spearman correlation coefficient. Non-parametric Wilcoxon signed-rank test was applied to evaluate the differences in CS measurements between the two tests and the agreement was assessed by Bland-Altman analysis in all groups of subjects. Linear regression analysis was used to evaluate the relationship between age and distance and near CS tests. Kruskal-Wallis H test was used to evaluate the ability of each CS test to discriminate between the three groups of subjects.

To examine the relationship between measures of vision (distance VA, near VA, distance CS and near CS) and reading performance metrics (RA, MRS and CPS), linear regression model was fit to investigate the prediction and association between measures of vision and reading performance metrics.

8.5.4 Repeatability of near visual acuity measurement with different optotypes

For all subjects (young, presbyopic and cataract), near visual acuity was measured with logarithmic near visual acuity chart (Precision Vision) with different optotypes. These optotypes were: Landolt-C, Tumbling-E, numbers and letters (ETDRS). Also, the reading acuity was measured using Bailey-Lovie word reading chart. All visual acuity and reading acuity were performed monocularly at distance of 40cm. The charts were viewed in a randomized order and presented under the same illumination condition. The subjects were instructed to read the near acuity chart until no more optotype was seen correctly. Each optotype has a value of 0.02 logMAR and single letter scoring method was used to identify the near VA. Subjects were given around a three-minute break between measures. All subjects were examined with best near correction, if needed. Two to three weeks later, the participants repeated the same near visual acuity tests.

The data in this experiment were normally distributed (Shapiro-Wilks >0.05). The test-retest repeatability was assessed using Pearson correlation coefficient and the difference of near acuity measurement between the two testing sessions was analysed for significance by mean of the paired t-test. To assess the agreement between the two sessions, Bland-Altman analysis was used to calculate the limit of agreements (LoA) as ± 1.96 SD of the difference, CI at 95%. Inter-chart reliability of the near VA charts was determined by calculating the Cronbach's alpha coefficient. Each near VA optotype and reading acuity was compared with each other in pair-wise manner by calculating the Pearson correlation coefficient and paired sample t-test was applied to evaluate the difference between the two measurements in each pairwise. Two-sided p value <0.05 was considered to be statistically significant.

8.6 Results

8.6.1 The accuracy of reading speed measurement by stopwatch versus audio recording method

The mean reading time measured by stopwatch (4.16 ± 8.6 seconds) was statistically significant longer than the mean reading time measured using the audio recording (3.83 ± 0.37 seconds). This difference in the reading time was in seconds: 0.33 ± 0.38, 95% confidence interval CI: (0.23; 0.43), $p < 0.001$. Of the 116 measurements, 84.48% of the stopwatch measurements were longer than the corresponding audio recording measurements, 12.06% of the stopwatch measurement were shorter than audio recording measurements and only

3.44% did not differ. Bland-Altman analysis found a low agreement for measuring the reading time of the sentences between stopwatch and audio recording methods (Fig 8.2a).

Audio recording measurement revealed significantly more reading errors than that recorded manually using the scoring sheet (median difference = 3 words, p <0.001). This difference in number of reading errors between the two recording methods influences the MRS measurement because the reading speed is calculated as the number of words read correctly divided by the reading time. The Wilcoxon signed-rank test found small but statistically significant difference in MRS measurement (median difference = 4 wpm). The stopwatch recording method resulted in slower reading speed compared to the audio recording. However, the Bland-Altman agreement was high for MRS measurements (Fig 8.2b).

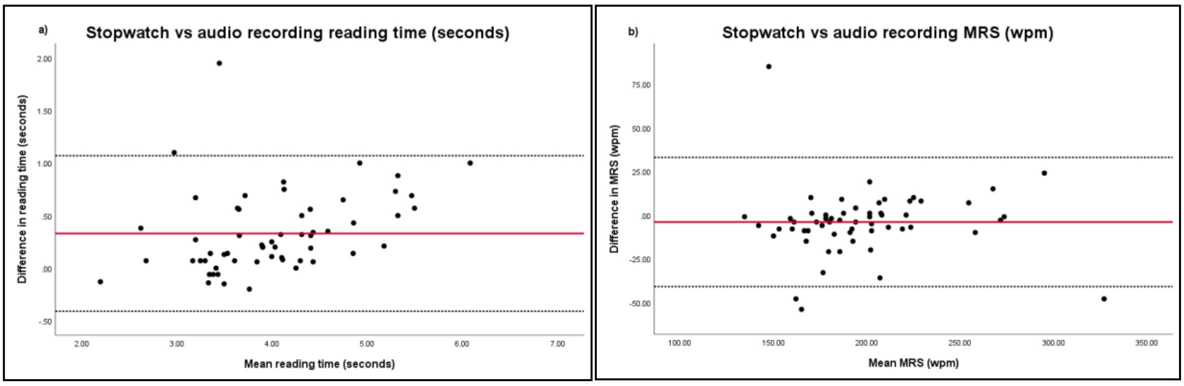


Figure 8.2: Stopwatch measurements compared with audio recording measurements. (a) the Bland-Altman analysis plot showed a low agreement between these two recording methods for measuring the MNread sentences reading time in seconds; (b) the agreement was high for measuring MRS (wpm).

8.6.2 Eye gaze position and reading speed

The results showed a strong correlation in the reading speed measurements between reading in an armchair and reading at a desk in young (r_s = 0.865) and in older subjects (r = 0.892). Also, no statistically significant differences were found in the reading speed results between the two reading conditions in both groups p >0.05.

The number of reading errors appeared to be unaffected by reading the sentences in different positions as no significant differences were found. Overall, read in armchair or read at desk conditions did not affect the reading performance results.

8.6.3 Correlation of measures of vision and reading performance

Table 8.1 compares the difference between distance CS measured with LCD test at 4m and near CS measured with Mars chart at 50cm. The Spearman correlation between the two tests revealed weak correlation in the young group of subjects (r_s = 0.324), moderate correlation in presbyopic subjects (r_s = 0.508) and better correlation found in cataract subjects (r_s = 0.723). In young and presbyopic subjects, near CS measurements were significantly lower than the distance CS test. The median differences in young and presbyopic subjects were 0.15 log units and 0.08 log units, respectively (Wilcoxon paired test p <0.05). The Bland-Altman analysis showed a low agreement between the distance and near CS tests in all groups of subjects (Fig. 8.3).

Linear regression analysis showed that there was a significant reduction in CS measurement with age in both distance [R^2 = 0.107, p = 0.007] and near [R^2= 0.107, p = 0.012] CS tests. This analysis was performed only in normally sighted young and older subjects to eliminate the influence of variation that may be caused by the presence of cataract. Interestingly, both CS tests showed that the age accounted for 10.7% of the variation in the measurements. Distance CS predicted a decrease of 1.87 log units per decade with upper and lower prediction interval 1.93 and 1.80 log units, respectively, while near CS test predicted a decrease of 1.74 log units per decade with upper and lower prediction interval 1.77 and 1.70 log units, respectively. Kruskal-Wallis H test showed that the difference between groups was statistically significant with both CS tests p <0.001.

Table 8.1: Difference between distance CS and near CS tests in the three groups of subjects. (r_s) Spearman's correlation coefficient, p-value of Wilcoxon signed-rank test and Bland-Altman agreement between two CS tests.

Subjects	r_s	***p* value**	**Median difference**	**Agreement**	**95% CI**
Young	0.324	<0.001	0.15	Low	(0.07, 0.15)
Presbyopic	0.723	0.002	0.08	Low	(0.03, 012)
Cataract	0.508	0.303	-0.03	Low	(-0.10, 0.02)

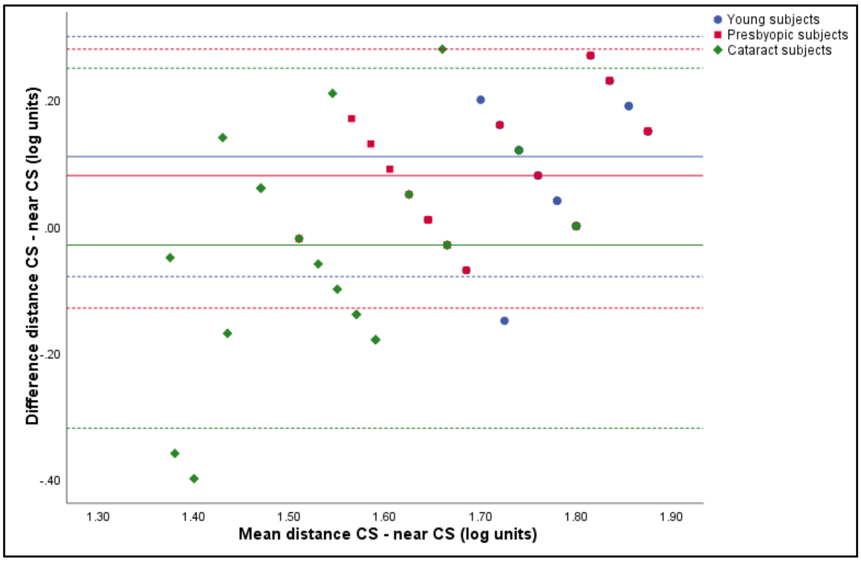


Figure 8.3: Bland-Altman plots of the differences between distance CS and near CS plotted against the mean of the two tests in three groups of subjects. The agreements between the two tests were low in all subjects.

The 84 subjects tested showed a relatively wider range of distance VA, near VA, distance CS and near CS in the cataract group of subjects compared to normally sighted subjects. The maximum, minimum, mean and standard deviation values for each measure of vision in the three groups of subjects are summarized below in Table 8.2.

Table 8.2: Measures of vision in the three groups of subjects

Measures of vision	Young		Presbyope		Cataract	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Distance VA (logMAR)	-0.10 (0.08)	0.1,-0.2	-0.07 (0.08)	0.1,-0.2	0.11 (0.14)	0.34,-0.1
Near VA (logMAR)	-0.06 (0.08)	0.08,-0.24	-0.01 (0.08)	0.14,-0.18	0.23 (0.14)	0.6,-0.04
Distance CS (log units)	1.84 (0.10)	1.95,1.65	1.75 (0.16)	1.95,1.35	1.55 (0.16)	1.80,1.20
Near CS (log units)	1.72 (0.07)	1.80,1.48	1.67 (0.09)	1.80,1.48	1.58 (0.11)	1.80,1.36

The association and the ability of each measure of vision to predict the reading performance metrics are shown in Table 8.3. The relations of the linear regression model only fit when the significance of the coefficient of determination is at 0.05 level. Distance visual acuity associated with CPS and reading acuity measurements in young subjects and not associated with any of the reading performance metrics in presbyopic and cataract subjects. Near visual acuity measured with ETDRS letter chart strongly associated with reading acuity measured with Bailey-Lovie chart in young and cataract subjects, but not in presbyopic subjects, and was responsible for 33.1% and 22.6% of the variation, respectively. Also, the near visual acuity has ability to predict CPS in cataract subjects only. Distance contrast sensitivity associated with RA and CPS in young group of subjects. However, distance CS was not associated with any of the reading performance metrics in presbyopic and cataract subjects. Near contrast sensitivity has a strong relationship with reading acuity (was predicted as 25% of the variation) and CPS measurements (was predicted as 11.8% of the variation) in cataract subjects. Also, near contrast sensitivity significantly associated with RA in young subjects. None of the measures of the vision had the ability to predict the reading speed in all groups of subjects.

Table 8.3: Associations between the measures of vision and reading performance metrics for the Bailey-Lovie word reading chart in young, presbyopic and cataract subjects. * linear regression model is significant at 0.05 level

Subjects	Reading metrics	Measures of vision	Adjusted R^2	Significance for coefficient of determination
Young	RA	Distance VA	0.286	0.002*
		Near VA	0.331	0.001*
		Distance CS	0.133	0.029*
		Near CS	0.083	0.041*
	MRS	Distance VA	-0.01	0.401
		Near VA	-0.035	0.808
		Distance CS	0.001	0.322
		Near CS	-0.008	0.388
	CPS	Distance VA	0.190	0.010*
		Near VA	0.028	0.190
		Distance CS	0.425	0.000*
		Near CS	0.071	0.087
Presbyope	RA	Distance VA	-0.007	0.378
		Near VA	0.040	0.153
		Distance CS	0.028	0.192
		Near CS	-0.001	0.333
	MRS	Distance VA	-0.030	0.661
		Near VA	-0.004	0.883
		Distance CS	-0.035	0.834
		Near CS	-0.008	0.383
	CPS	Distance VA	0.097	0.066
		Near VA	-0.020	0.504
		Distance CS	-0.037	0.926
		Near CS	-0.032	0.727
Cataract	RA	Distance VA	-0.061	0.444
		Near VA	0.226	0.008*
		Distance CS	-0.039	0.820
		Near CS	0.250	0.005*
	MRS	Distance VA	-0.040	0.829
		Near VA	-0.032	0.644
		Distance CS	-0.015	0.431
		Near CS	0.013	0.262
	CPS	Distance VA	-0.027	0.560
		Near VA	0.114	0.041*
		Distance CS	-0.042	0.994
		Near CS	0.118	0.048*

8.6.4 Repeatability of near visual acuity measurement with different optotypes

The results of the repeatability of the near VA with different logMAR optotypes in the three groups of subjects are summarized in Table 8.4. Landolt-C, Tumbling-E, numbers and ETDRS revealed good Spearman correlation coefficients and high agreements without any statistical significance between the two sessions in cataract subjects. Landolt-C showed a statistically significant difference in young subjects and this difference was logMAR: 0.03 ± 0.05, 95% confidence interval CI: (0.009; -0.054), p = 0.008. The repeatability results of the Landolt-C in presbyopic subjects were poor and showed a statistically significant difference logMAR: 0.03 ± 0.08, 95% CI (0.004; 0.07), p = 0.026 with moderate correlation (r = 0.570) and low agreement between the two testing sessions. Although the correlation and agreement of Tumbling-E optotype were high in young and presbyopic subjects, significant differences were found. These differences in young and presbyopic subjects were logMAR: 0.02 ± 0.06, 95% CI (0.003; 0.052), p = 0.028 and logMAR: 0.04 ± 0.06, 95% CI (0.02; 0.07), p = 0.002, respectively. Numbers optotypes revealed a good correlation and agreement in young and presbyopic subjects. However, a statistically significant difference between the two sessions was found in young subjects. The repeatability results of the ETDRS showed a significant difference in young and presbyopic subjects and these differences were logMAR: 0.05 ± 0.18, 95% CI (-0.01; 0.12) p = 0.004 and logMAR: 0.04 ± 0.06, 95% CI (0.018; 0.07) p = 0.001, respectively. The 95% LoA of ETDRS was low in young subjects while it was high for presbyopic subjects. Overall, the Bland-Altman agreements were high for all the optotypes in all subjects except the Landolt-C in presbyopic subjects and the ETDRS in young subjects. The Cronbach’s alpha was high for the four optotypes and was above 0.7, which indicates good inter-chart reliability results.

Table 8.4: Test-retest repeatability of the near VA with different optotypes in three groups of subjects. T-test of the differences between the two testing sessions is significant at 0.05 level

Group	Near VA optotypes	Correlation coefficient (r value)	Differences in measurements (t-test) *p* value	95% CI (LoA)	95% CR
Young	Landolt-C	0.756	0.008	[0.14; -0.08]	0.29
	Tumbling-E	0.708	0.028	[0.13; -0.09]	0.12
	Numbers	0.809	0.018	[0.11; -0.07]	0.11
	ETDRS	0.762	0.004	[0.18; -0.10]	0.19
Presbyope	Landolt-C	0.570	0.026	[0.18; -0.13]	0.17
	Tumbling-E	0.757	0.002	[0.15; -0.07]	0.13
	Numbers	0.722	0.826	[0.11; -0.11]	0.13
	ETDRS	0.692	0.001	[0.15; -0.07]	0.14
Cataract	Landolt-C	0.788	0.212	[-0.26; -0.20]	0.24
	Tumbling-E	0.645	0.123	[0.22; -0.16]	0.21
	Numbers	0.555	0.149	[0.31; -0.23]	0.28
	ETDRS	0.780	0.276	[0.19; -0.15]	0.20

Table 8.5: Measures of near visual acuity with different logMAR optotypes and reading acuity (RA) measured with Bailey-Love reading chart in all subjects.

			Mean ± SD		
Group	**Landolt-C**	**Tumbling-E**	**Numbers**	**ETDRS**	**RA**
Young	-0.05 ± 0.08	0.03 ± 0.08	-0.07 ± 0.10	-0.06 ± 0.08	-0.10 ± 0.10
Presbyope	0.04 ± 0.07	0.04 ± 0.07	-0.02 ± 0.09	-0.01 ± 0.08	-0.06 ± 0.09
Cataract	0.34 ± 0.18	0.27 ± 0.13	0.25 ± 0.15	0.23 ± 0.14	0.18 ± 0.13

The mean ± SD of the reading acuity and near visual acuity measures with different logMAR optotypes are presented in Table 8.5. From the data in Tables 8.6, 8.7 and 8.8 of pair-wise comparison between near VA and reading acuity in the three groups of subjects, it can be seen that no significant differences were found between Landolt-C and Tumbling-E in young and presbyopic subjects. However, the difference was significant in cataract subjects. The numbers optotypes resulted in significantly better near VA compared to Landolt-C and Tumbling-E. On the other hand, numbers optotypes correlated well with ETDRS (letters) without any significant differences in all groups of subjects. ETDRS always results in significantly better near VA compared to Landolt-C and Tumbling-E. The most interesting aspect in this analysis is the association between the reading acuity measured with the reading chart and near VA measured with different optotypes. Interestingly, near VA measured with Londolt-C, Tumbling-E, numbers and ETDRS poorly correlated with reading acuity. The results showed that the reading acuity always resulted in a better VA compared with other optotypes in all groups of subjects. Also, the Bland-Altman plots (Fig 8.4. 8.5 and 8.6) showed a low agreements between the near VA and reading acuity.

Table 8.6: Pearson correlation coefficients (r value) and t-test for significant differences (p value) for each pair-wise comparison between near visual acuity optotypes in young normally sighted subjects * the difference in near VA between two optotypes is significant at 0.05 level

Optotypes	**Landolt-C**	**Tumbling-E**	**Numbers**	**ETDRS**
Landolt-C	-	-	-	-
Tumbling-E	*r* = 0.572 *p* = 0.301	-	-	-
Numbers	*r* = 0.838 *p* = 0.022*	*r* = 0.642 *p* = 0.011*	-	-
ETDRS	*r* = 0.712 *p* = 0.031*	*r* = 0.637 *p* = 0.046*	*r* = 0.833 *p* = 0.216	-
RA	*r* = 0.411 *p* = 0.017*	*r* = 0.369 *p* = 0.003*	*r* = 0.607 *p* = 0.035	*r* = 0.604 *p* = 0.025*

Table 8.7: Pearson correlation coefficients (r value) and t-test for significant differences (p value) for each pair-wise comparison between near visual acuity optotypes in presbyopic subjects * the difference in near VA between two optotypes is significant at 0.05 level

Optotypes	**Landolt-C**	**Tumbling-E**	**Numbers**	**ETDRS**
Landolt-C	-	-	-	-
Tumbling-E	*r* = 0.603 *p* = 0.667	-	-	-
Numbers	*r* = 0.729 *p* <0.001*	*r* = 0.644 *p* <0.001*	-	-
ETDRS	*r* = 0.766 *p* <0.001*	*r* = 0.592 *p* <0.001*	*r* = 0.719 *p* = 0.263	-
RA	*r* = 0.452 *p* <0.001*	*r* = 0.555 *p* <0.001*	*r* = 0.579 *p* = 0.011*	*r* = 0.530 *p* = 0.002*

Table 8.8: Pearson correlation coefficients (r value) and t-test for significant differences (p value) for each pair-wise comparison between near visual acuity optotypes in cataract subjects * the difference in near VA between two optotypes is significant at 0.05 level

Optotypes	**Landolt-C**	**Tumbling-E**	**Numbers**	**ETDRS**
Landolt-C	-	-	-	-
Tumbling-E	*r* = 0.872 *p* = 0.001*	-	-	-
Numbers	*r* = 0.709 *p* = 0.001*	*r* = 0.706 *p* = 0.332	-	-
ETDRS	*r* = 0.801 *p* <0.001*	*r* = 0.775 *p*= 0.022*	*r* = 0.812 *p* = 0.209	-
RA	*r* = 0.648 *p* <0.001*	*r* = 0.508 *p* = 0.002*	*r* = 0.446 *p* = 0.034*	*r* = 0.455 *p* = 0.020*

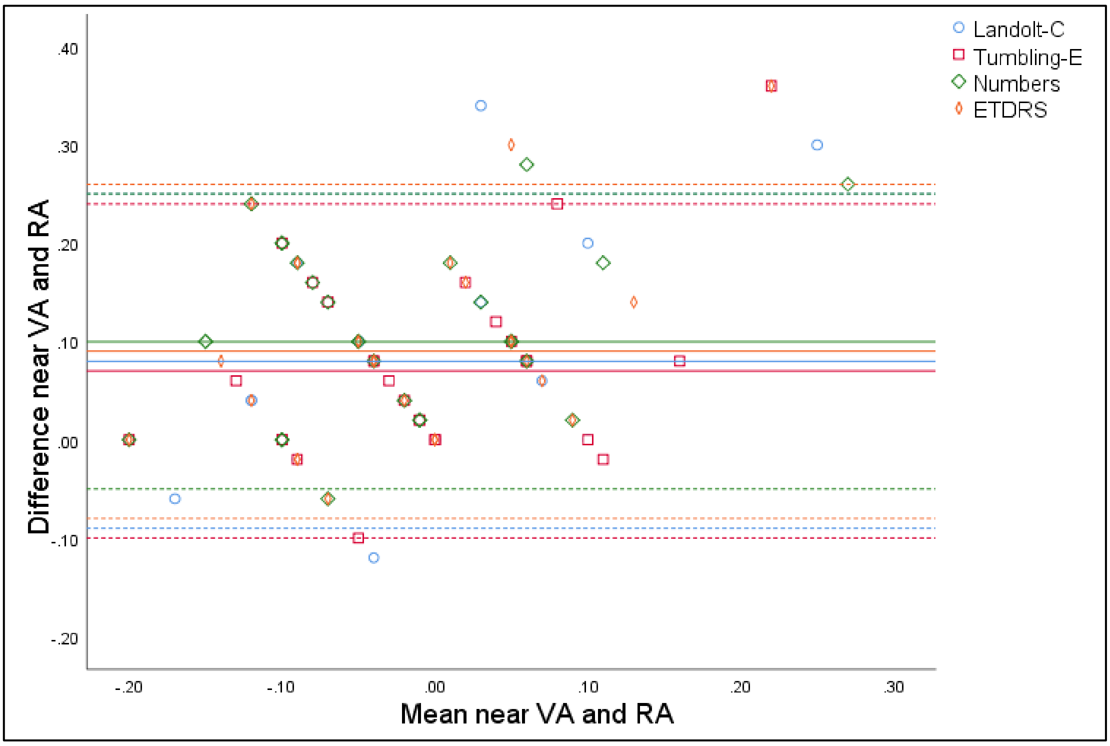


Figure 8.4: Comparison of near VA measured with Landolt-C, Tumbling-E, numbers and ETDRS optotypes with reading acuity in young normally sighted subjects. Bland-Altman plots showed a low agreements between near VA and reading acuity for all optotypes except for the number optotype, the agreement was low.

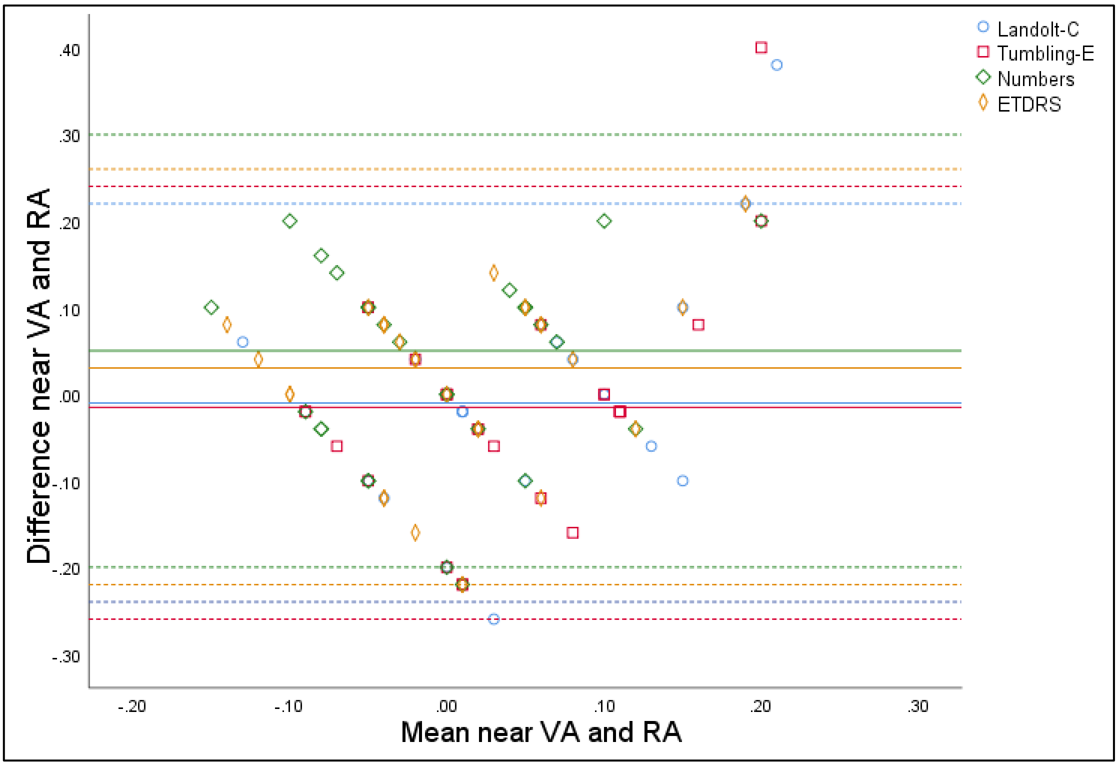


Figure 8.5: Comparison of near VA measured with Landolt-C, Tumbling-E, numbers and ETDRS optotypes with reading acuity in presbyopic subjects. Bland-Altman plots showed a low agreements between near VA and reading acuity for all optotypes except for the number optotype, the agreement was high.

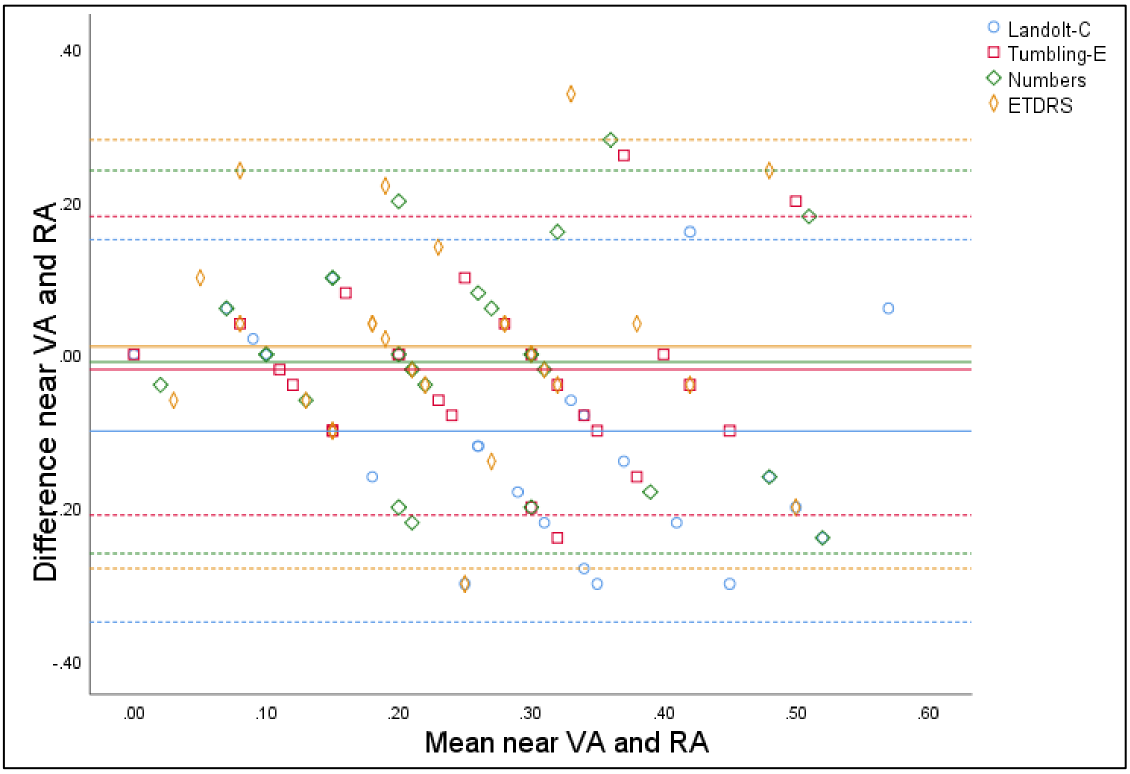


Figure 8.6: Comparison of near VA measured with Landolt-C, Tumbling-E, numbers and ETDRS optotypes with reading acuity in cataract subjects. Bland-Altman plots showed a low agreements between near VA and reading acuity for all optotypes.

8.7 Discussion

8.7.1 The accuracy of reading speed measurement by stopwatch versus audio recording method

The present study shows that the stopwatch measurements of reading time and reading speed differ from those of an audio recording method. Although the reading time measured manually with stopwatch is easy and can be calculated during the test, the measurement can be influenced by the examiner's reaction time. While recording the time at the end of the sentence can be easily determined by the examiner, determining the onset of reading, however, more depends on the examiner's reaction time; thus, it could be a potential source of the variability in the measurements. The results showed that the difference in the MRS measurements between the two recording methods was found to be within acceptable limit as it was only four words per minutes.

Reading test charts such as MNread, Radner and Bailey-Lovie use short sentences with a reading time ranged from three to six seconds. It has been found that the variability of the reading time measurements increased with shorter sentences compared to the longer text paragraph (Brussee et al., 2015). Therefore, in a short sentence reading test such as usually used in clinical practice or for research, it is important to control any obvious source of the variability, such as examiner reaction time on manual stopping of the stopwatch.

The results showed that more reading errors were revealed when analysing the data afterward using the audio recordings compared to stopwatch method. This result is not surprising, because, when the stopwatch method is used, the examiner has to not only watch the score sheet for possible reading errors, but also observe the reader to determine the onset of the reading time and ensure that the reader is following the correct procedure. In addition, the examiner also has to watch and control the stopwatch. These synchronized demands on the examiner are possible reasons to miss some of the words read incorrectly and are likely to increase the variability of the reading performance results.

Using a stopwatch to measure the reading time of each sentence in a reading chart in order to calculate the reading speed is known as reliable and well-accepted method (Bailey and Lovie, 1980, Mansfield et al., 1993, Ahn et al., 1995, Ahn and Ledge, 1995). The finding of this study has important implications for comparing with studies that used different recording methods to measure the reading time. For example, Subramanian and Pardhan (2009) and Patel et al. (2011) both investigated the test-retest reliability of the MNread chart in visually impaired subjects. The first study used the audio recording method and reported much better repeatability with 95% CR of 0.2 logMAR for CPS and 0.1 logRAD for RA, compared to the second study which used the stopwatch method and reported 95% CR of 0.55 logMAR for CPS and 0.3 logRAD for RA. The audiotape measurement in the study by Subramanian and Pardhan (2009) might have reduced the variability and, therefore, led to higher repeatability results. Therefore, this case supports the view that making comparison about the characteristics of the reading charts is rather difficult. This thesis attempt to control all the measurement-related factors that may contribute to the variability in order to evaluate the repeatability and reliability of the reading test charts more precisely.

Using the stopwatch and the scoring sheet for reading errors at the same time may become a source of variability in the reading performance measurement. Although digital audiotaping or videotaping afford an attempt to overcome this problem, it is not usually feasible to use these in clinical practice. The solution may be addressed by using a computerized reading speed test that measures the reading time automatically, such as the Radner app discussed in the previous chapter.

8.7.2 Eye gaze position and reading speed

This experiment is designed to examine the reading speed of young and older subjects under different reading conditions (straight ahead and head downward) to establish and understand the best methodology for measuring the reading performance in research studies or clinical practices.

This experiment presents a simple picture of the clinical settings most frequently used during the reading performance test. The results showed that the reading speed and reading error measurements are not influenced by the two reading positions. However, in read in armchair position, it is important to constantly control the reading distance because reading the smaller print sizes may be accompanied by sharp downward head turn, which may lead to decrease the viewing distance (Marumoto et al., 1998). It has been found that small change in the reading distance may lead to different test results (Patel et al., 2008).

Since the reading performance measurements are not affected by reading positions, knowing and examining the actual behaviour of the reader may be crucial in enhancing the reading efficiency and reader's comfort.

8.7.3 Correlation of measures of vision and reading performance

The first aim of this section was to evaluate the differences and the agreements between distance CS measured with LCD system and near CS measured with Mars Letter CS test in the young, presbyopic and cataract subjects. LCD CS test and Mars chart have been evaluated and compared to the standard Pelli-Robson test, but no published data have compared the LCD CS test with the Mars chart.

The two CS tests have shown the ability to discriminate between the three groups of subjects. Also, it was generally accepted that contrast sensitivity decreases with age. The age of the normally sighted subjects was from 18 to 78 years and both distance and near CS tests showed that age significantly predicts a decrease of contrast sensitivity. This confirms that each CS test is reasonable to examine contrast sensitivity. However, significant differences and low Bland-Altman agreements were found between the near and distance CS tests in all groups of subjects. Near CS resulted in lower CS compared to the distance CS test. These differences were 0.15 log units in young subjects, 0.08 log units in presbyopic subjects, while the difference was smaller for cataract subjects at 0.03 log units. These results are in agreement with those obtained by Haymes et al. (2006) who stated that the Mars chart resulted in worse CS in individuals with good CS, and vice versa. It seems possible that these results are because each letter in LCD CS corresponds to 0.05 log units, while corresponding to 0.04 log units in the Mars CS test. Another possible explanation is the higher luminance of the LCD monitor compared to the Mars printed chart. Thus, these results provide insight to practitioners to be aware of using the same CS test for comparing the outcomes of the ocular treatment.

The second objective of this section was to examine the association between distance VA, near VA, distance CS and near CS and the reading performance metrics assessed by Baily-Lovei chart. An implication of this analysis is the possibility that practitioners need to estimate the reading performance of the patient from the primary clinical data without performing the full reading assessment.

The results showed that distance VA and distance CS tests played an important role in prediction of RA and CPS in young participants only. However, these results are not very encouraging because these two tests do not have the ability to predict the reading performance metrics in presbyopic and cataract subjects, where they showed a wider range the measurements compared to the young normally sighted subjects. This finding provides a further support to studies that have found the distance visual acuity is a poor predictor for reading performance (Ahn and Ledge, 1995, Stifter et al., 2005a). The results of the poor association between distance CS and reading performance in cataract and presbyopic subjects need to be interpreted with caution. As the distance CS was measured with an LCD monitor, the possibility is that the presentation of the low level of contrast may be suboptimal due to the chart high luminance. It has been shown that the LCD monitors were less suitable to measure the contrast sensitivity in visual psychophysical studies than cathode ray tube displays (Strasburger et al., 2002). Another source of uncertainly is that, in contrast to the Mars CS test, the LCD CS showed a poor agreement with standard Pelli-Robson chart (Thayaparan et al., 2007, Zeri et al., 2018).

The prediction of the reading performance in people with poorer contrast sensitivity, such as in cataract subjects, is more favourable. It was noteworthy that the near visual acuity and near CS associated better with reading acuity and CPS measurements in cataract subjects. Regression analysis showed that the near VA and near CS were responsible for 22.6% and 25%, respectively, of the variation in reading acuity, and for 11.4% and 11.8% of the variation in CPS, respectively. The near CS test played a more important role in detection of reading performance than distance CS tests because it was measured with Mars CS test, which has been proven to have reliability results equal to or better than Pelli-Robson test (Dougherty et al., 2005, Arditi, 2005, Haymes et al., 2006). In addition, there are many characteristics that make the Mars test suitable to use in clinical practice as it is small, portable, measured at near distances, easier to administer and illuminates equally. The finding of the strong correlation between contrast sensitivity loss and reading performance was also reported by (Brussee et al., 2018, Mones and Rubin, 2005) and Brussee et al. (2018), Thus, these association between near CS and reading performance metrics (reading acuity and CPS) provide the rationale for including a valuable CS test in clinical practice as the primary measure to assess the treatment outcome in addition to a visual acuity test. It should be noted that none of the measures of vision had the ability to predict the reading speed in all groups of subjects. Therefore, the estimation of the reading speed needs a direct measure by using a reliable reading test chart.

8.7.4 Repeatability of near visual acuity measurement with different optotypes

The measurement of near visual acuity is widely used to evaluate the visual function because of its simplicity. However, different types of optotype can be used in clinical practice. The findings of this study provide evidence that evaluating the same subject with different near optotypes resulted in different near visual acuity outcomes.

As was pointed out in Chapter 2, reading difficulty was one of the common reasons to visit the optometric practice, therefore suggesting patients were more concerned about their reading ability. It has been addressed that, during the near vision assessment, the ability to resolve the word was more realistic and represented the patient’s daily life to a greater extent as opposed to than isolated near optotypes (Mansfield et al., 1993). This suggestion reflects those of Xiong et al. (2018) who also found that reading acuity is a better predictor than letter acuity for fluent reading and spot reading. The results showed that the reading acuity measurements did not agree and differed significantly from Landolt-C, Tumbling-E, numbers and ETDRS near optotypes. Reading acuity always resulted in better VA compared to other near optotypes in all groups of subjects. This finding is contrary to previous studies which have suggested that the word optotype resulted in a worse visual acuity compared to single-letter optotype (Gupta et al., 2009b, Sheedy et al., 2005). However, the results of better reading acuity compared to the isolated optotypes are consistent with those of Chung et al. (2007) and Xu and Bradley (2015) who used the continuous-text MNread chart and associated the reason for the better reading acuity measurements to the context benefit of the meaningful sentences. This effect was absent in this study because the context benefit was offset by using the unrelated word Bailey-Lovie chart. The extent of the crowding effect may explain the relationship between reading acuity and letter acuity (Bouma, 1971).

The advantage or reading acuity over near visual acuity may depend on parafoveal-preview benefits. The readers obtain information from parafoveal vision in some fashion, e.g., skipping pattern when the information obtained is not only from the fixated word (foveal), but also from the following word ‘parafoveal’ (Blanchard et al., 1989). There has been a growing interest in which parafoveal words affect fixation time. It has been found that if the parafoveal information is restricted, the reading performance decrease quickly (Rayner et al., 1982), giving that advantage of reading acuity over letter acuity. Some studies confirmed that meaningful sentences are skipped more frequently than unrelated words sentence (Rayner et al., 1996). However, later on, it has been proven that the landing position is a little influence by contextual constrain (Rayner et al., 2001). This is in the line of this study since the reading acuity measured with unrelated words Bailey-Lovie chart.

The near VA measured with Landolt-C and Tumbling-E did not differ significantly in normally sighted subjects, but showed a significant difference in subjects with cataract. Landolt-C resulted in worse VA by approximately one whole line compared to the Tumbling-E. Grimm et al. (1994) reported that the Tumbling ‘E’ optotype was easier to recognise than Landolt rings optotype, as the two gaps in the letter E, in comparison to the one gap in the C optotype, enhanced better recognition of Tumbling ‘E’.

Overall, the test-retest Spearman correlation coefficients for the near optotypes revealed a relatively strong correlation range from 0.6 to 0.8. However, Landolt-C, Tumbling-E and ETDRS showed a significant difference between the two testing sessions in young and presbyopic subjects. The differences in the repeated measurement in Landolt-C and

Tumbling-E may be due to that, when the smaller ‘C’ and ‘E’ optotypes are presented in one line, the patient may confuse the examiner by reading backwards, not starting at the beginning of the line or losing their place, especially in that reading C and E in different directions is not familiar to the patient as in the case of letters or numbers. Therefore, asking the participant to repeat the response may produce a source of errors in the repeated measurement. Although, Landolt-C is the standard reference optotype in vision testing for use in the laboratory and research studies, it is not recommended to use in clinical practice because it is tends to be poor for general clinical application (NAS-NRC, 1980).

The visual characteristics of Landolt-C and Tumbling-E demanded of the person’s visual system is to identify the directions and the shapes (resolution acuity), compared to reading individual numbers and letters (recognition acuity). The ETDRS chart has become the most frequently used optotype in clinical setting for recognition acuity (Iii et al., 1993). Generally, resolution and recognition acuities are supposed to be equivalent in their ability to assess visual acuity. However, the results showed that ETDRS and number optotypes have always resulted in a better near VA compared to Landolt-C and Tumbling-E. In order to explain why recognition acuity is better than resolution acuity, the properties of the optotypes should be considered. It has been established that the use of the letters components such as curves or diagonals facilitate the accurate identification of the letters, even if the entire letter is not clear. Such component cues are absent in resolving the gap in E or ring, thereby explaining the worse acuities (Wittich et al., 2006). In addition, it has been reported that the letter C which is included in ETDRS chart is more difficult to recognize than other letters (Sheedy et al., 1984, Iii et al., 1993).

It seems that the (number) optotype has a high standard when assessing the near visual acuity in terms of repeatability and reliability. In number optotype, good correlation and no significant differences were found between the two testing sessions in presbyopic and cataract subjects. Although statistically significant difference was found in young subjects, this difference was clinically not significant as it was very small and approximately equal to one letter difference (0.03 logMAR). In addition, the (number) optotype VA correlates well with ETDRS in all groups of subjects. Some of the advantages for using the number optotype are the familiarity and minimal instruction required for the patients.

8.8 Conclusion

The findings of this chapter have several important implications in both clinical practice and research studies:

1. Studies on the reliability of the reading test charts need to be done under a standardized condition. This study has identified that the reading time and the reading errors results differ when measured with stopwatch and scoring sheet to those measured after the test completion using the audio recordings.
2. The negative results about the effect of the different reading positions on reading speed measurements should encourage practitioners to test the reading performance in a more relaxed sitting for the patient while controlling the reading distance.
3. Practitioners should consider that the computer-based LCD CS test is not interchangeable with the printed Mars Letter CS test. However, both tests have shown a reasonable ability to examine and differentiate the CS results between normally sighted subjects compared to subjects with cataract.
4. In a clinical setting there is a time restriction when attempting to fully assess the individual's reading performance. Thus, it would be useful to determine which primary visual outcome is most relevant to reading performance. The results showed that no single measure of vision has the ability to predict the variability of the reading performance metrics sufficiently. However, including the Mars CS test in addition to near visual acuity may provide a good understanding of the reading performance, especially for people with poorer CS, such as cataract patients.
5. The difference in near visual acuity measurement between different optotypes suggests that the redundancy of measuring near VA should not be performed with more than one type of optotype. Number near optotype resulted in a good repeatability results for measuring near VA in both young and older normally sighted subjects and in subjects with cataract. The results showed that the reading acuity did not agree with near VA measured with isolated optotype. Therefore, measuring the reading acuity in clinical practice may have more important value and provide more accurate assessment because it is more representative of real-word reading performance.

Chapter 9

Conclusion

9 Introduction

The literature has many studies that have attempted to demonstrate the reliability of reading test charts. However, there is a variability between the studies in the measurement setting, statistical approach and the text language used. Thus, the agreement and the reliability estimates are confounded by these sources of variation (Kottner et al., 2011).

The first experimental chapter (Chapter 2) of this thesis confirmed that reading difficulty is the primary reason for visits to optometric practices irrespective of age or gender. Thus, if the reading performance measurement is to continue to increase as a critical clinical outcome of eye disease and treatment intervention, there must be agreement on which reading test should be used. Many different unstandardized reading charts are currently used in clinical practices among the optometrists as highlighted in Chapter 2.

An important consideration regarding which reading test chart to choose is whether it has been tested and developed according to scientific standards. There are certain well accepted standards for selection and comparison between tests (Rubin, 2013). These are based on evaluation of reliability and repeatability. Reliability means (does the test measure exactly what it is intended to measure?). While the repeatability refers to (are the results of the test repeatable and consistent?). Evaluation of the sensitivity and specificity is also needed for a test used for diagnostic purposes. Sensitivity of the test refers to the ability of the test to correctly identified the patients with ocular disease whilst the specificity refers to the ability of the test to correctly identified the people without disease.

This thesis attempts to evaluate the repeatability and reliability of different standardized reading test charts in one study on three groups of subjects. First, in young normally sighted subjects, because the evaluation of the reading test charts should be first calibrated among normally sighted people to eliminate any factors that may affect the reading performance measurements. Second, in presbyopic subjects to evaluate the reading test effectiveness in order to examine one of the most important concerns of presbyopic sufferers. Finally, in subjects affected by cataract to assist in understanding the properties of reading test chart when used by people with eye diseases. This is the first study to make a direct comparison of six different reading charts in the English language where measurement and patient related factors that may affect the results of the reading performance metrics, are controlled. The following sections will illustrate how each reading chart performs in terms of test repeatability, reliability, sensitivity and specificity with the three groups of subjects.

Before that, this thesis answered the question of whether the clinical results of the reading performance metrics vary between different reading test charts. Correlation and agreements analyses of the reading performance metrics confirmed that different reading performance outcomes were obtained when using different reading test charts on the same participant. This finding has important implications for choosing the same test chart for consecutive examinations of the same patient or for comparison in research.

9.1 Colenbrander reading chart

The Colenbrander reading chart showed low Bland-Altman agreement between the two testing sessions for measuring the reading acuity in the three groups of subjects. In young and presbyopic subjects, the correlations between the two sessions were moderate to low for measuring MRS and CPS. The coefficient of repeatability (CR) showed that the Colenbrander reading chart was less repeatable than MNread, Radner and Bailey-Lovie chart for measuring MRS and CPS in presbyopic and cataract subjects. However, the inter-chart reliability was relatively good, except for CPS measurement in the young group of subjects where it was poor ($\alpha = 0.456$). Due to the short words used in the Colenbrander reading chart, it always resulted in a faster reading speed compared to the other charts. Because the evaluation of the reading performance in particular was based upon the RA, the poor repeatability results in the three group of subjects is sufficient evidence to conclude that the Colenbrander reading chart is not a good choice to evaluate the reading performance metrics.

To further assess how well the reading speed measurements in the Colenbrander chart discriminate between presbyopic and cataract subjects, a receiver operating characteristics (ROC) curve was plotted for sensitivity versus (1- specificity). The area under the curve was calculated as 0.666, which indicates a poor ability of the Colenbrander reading chart to discriminate between these two groups of subjects (Fig. 9.1).

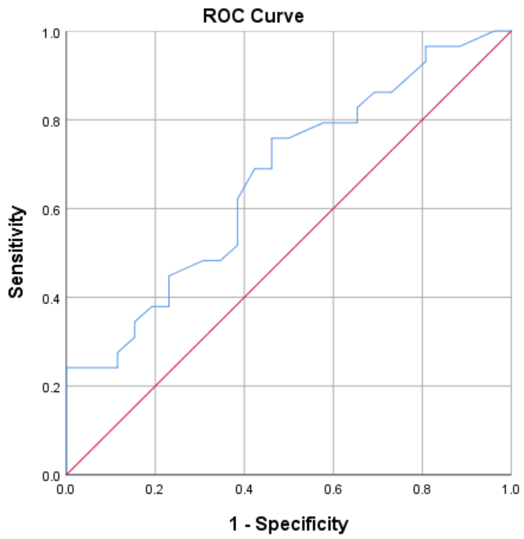


Figure 9.1: The ROC curve showing the ability of the Colenbrander reading speed to discriminate between patients with cataract and presbyopic subjects. Area under the curve equal to 0.666.

9.2 Radner reading chart

In young subjects, the Radner chart revealed a low agreement with statistically significant differences between the two testing sessions for measuring RA and MRS. Also, the correlations were low for measuring CPS and LogRAD/LogMAR ratio. In presbyopic subjects, the repeatability results showed statistically significant differences for measuring all the reading performance metrics. In multiple regression analysis, the coefficient of determination r^2 indicated that the cataract type and grade were responsible for 20% of the variation in the RA results rather than print sizes of the chart. As stated earlier, the print size should be the only parameter that affects the reading performance throughout the reading chart.

The high validity for the reading speed measurements of the reading test can be predicted by a good correlation with that obtained for long paragraphs (Trauzettel-Klosinski et al., 2012). The results showed that the 29 young normally sighted participants read the IReST significantly faster than the Radner (median difference= 32 wpm, p<0.05). However, the Radner reading speed has a fair ability to discriminate between cataract and presbyopic subjects, as the ROC was calculated as 0.703 (Fig 9.2).

While reviewing of the literature, it was noted that the Radner reading chart is the only chart that has been studied extensively in more than nine languages. However, when it was compared to the other charts in this study, the findings confirmed that, although the Radner chart is developed in highly standardized way, it is not the best choice for measuring the reading performance metrics.

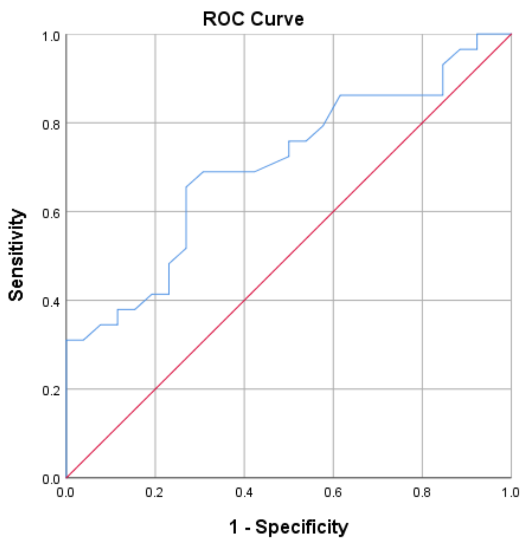


Figure 9.2: The ROC curve showing the ability of the Radner reading speed to discriminate between patients with cataract and presbyopic subjects. Area under the curve equal to 0.703.

9.3 The MNread reading chart

The repeatability results of the MNread chart did not show any statistically significant differences for measuring all the reading performance metrics in the three groups of subjects with good Bland-Altman agreements. Except for measuring logRAD/logMAR ratio in the prebyopic group only, the repeatability and the reliability (α= 0.492) were poor. The MNread inter-chart reliability was excellent for measuring the RA and MRS in all groups, Cronbach's alpha ranged from 0.883 to 0.921. The coefficient of repeatability for measuring MRS was less than the other charts (more repeatable) in cataract subjects while it was more (less repeatable) in presbyopic subjects. The young subjects read the long paragraphs in IReST significantly faster than the MNread (median difference = 19 wpm, p<0.05). However, this difference was less than that found in the Radner test.

In the MNread chart, the cataract type and grades accounted for 23.3% of the variation in RA results and 30.2% of the variation in the CPS results. ROC curve showed a good ability of the MNread chart to discriminate between cataract and presbyopic subjects and calculated as 0.721 (Fig 9.3). In general, the MNread chart can be used reliably to measure the reading performance metrics.

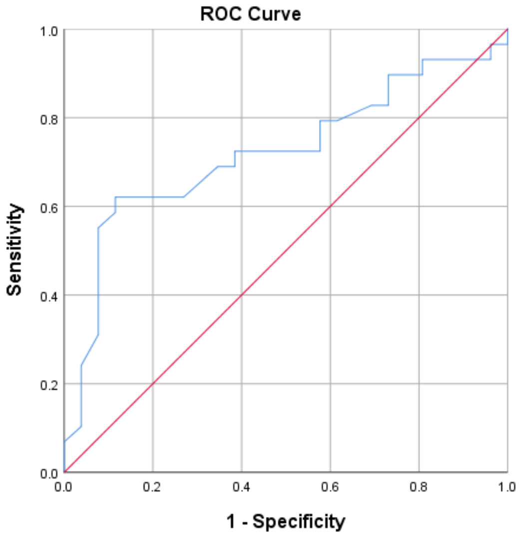


Figure 9.3: The ROC curve showing the ability of the MNread reading speed to discriminate between patients with cataract and presbyopic subjects. Area under the curve equal to 0.721.

9.4 The Bailey-Lovie unrelated words reading chart

This study is the first evaluate the Bailey-Lovie unrelated words reading chart characteristics in terms of reliability and repeatability. However, the results were very encouraging. The agreements between the two testing sessions were high for all the reading performance metrics without statistically significant differences in the three groups of subjects. Although, in young subjects, the difference in RA was significant, $p < 0.05$, it was clinically not significant as the difference was below one line of reading acuity. The inter-chart reliability results were high. The difference in reading speed between long paragraph IReST and Bailey-Lovie was 90 wpm slower, and this difference was bigger than that found in Colenbrander, Radner and MNread. This is because the Bailey-Lovie reading chart consists of unrelated random words rather than continuous text.

Interestingly, the Bailey-Lovie reading chart shows the minimal dependence on the cataract type and grade, as the coefficient of determination was small and less than 6% for measuring RA, MRS, CPS and LogRAD/LogMAR ratio. This indicates that the print size in the chart is the only parameter responsible for the variation of the results. The ROC curve showed a very good ability of the Bailey-Lovie reading speed measurements to discriminate between cataract and presbyopic subjects and calculated as 0.750 (Fig 9.4). This value was larger than that found in Colenbrander, Radner and MNread. Although, the Bailey-Lovie reading chart is the oldest one compared to the other reading charts, the finding indicate that it is the best chart to provide highly reliable and repeatable reading performance results in all groups of subjects.

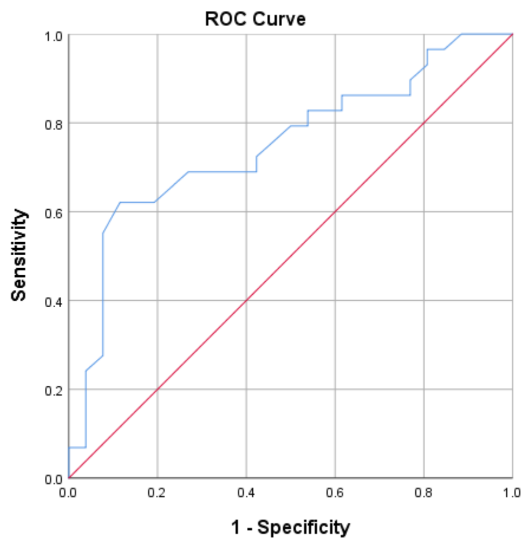


Figure 9.4: The ROC curve showing the ability of the Bailey-Lovie reading speed to discriminate between patients with cataract and presbyopic subjects. Area under the curve equal to 0.750.

9.5 IReST reading speed test

The IReST reading speed test was an excellent choice for measuring reading speed only as it has one print size. However, the IReST reading chart is not as versatile as the other reading charts because it cannot measure RA, CPS or logRAD/logMAR ratio. The agreements between the two testing sessions were high for all groups of subjects and Cronbach's alpha was above 0.9 in young and presbyopic subjects and equal to 0.886 in cataract subjects. The coefficient of repeatability results showed that the IReST was less repeatable in cataract subjects (CR= 11) than in presbyopia subjects (CR= 6). The different cataract types influenced the results of the reading speed in the IReST (28.6%). However, it has a very good ability to discriminate between cataract and presbyopic subjects as the ROC curve was calculated as 0.781 (Fig 9.5).

Therefore, this result suggests that the IReST is a valuable test for one specific aspect of reading in normally sighted subjects. However, due to small print size not matching the reading acuity level in subjects with cataract, the repeatability was reduced. In this case, it is better to measure the reading speed using a reading chart with graduated print sizes. It has been recommended to use visual aids in low vision subjects to compensate for the lower resolution (Trauzettel-Klosinski et al., 2012).

Since it is not feasible to measure the average reading speed from all 10 paragraphs in IReST during clinical examination or in research studies, this study evaluated the repeatability and reliability of each text paragraph in the IReST. The results showed a variation between the texts and revealed that the last five paragraphs had better repeatability results that the first five paragraphs. Therefore, a reliable paragraph must be chosen. This finding highlights the importance of investigating the reliability of every single sentence or paragraph in a reading test chart. The validity of a reading test is high when it has equally difficult sentences or paragraphs (Brussee et al., 2015).

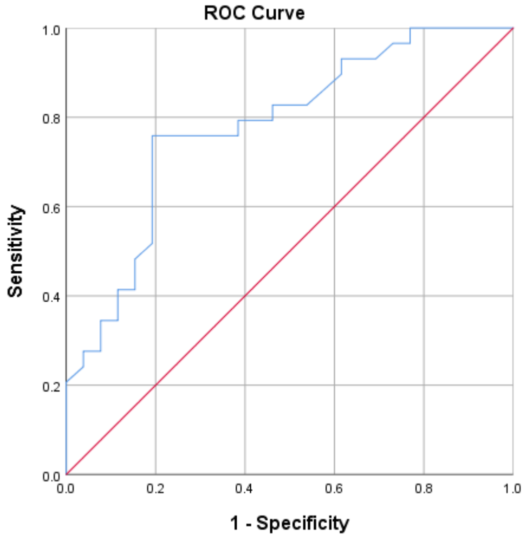


Figure 9.5: The ROC curve showing the ability of the IReST reading speed to discriminate between patients with cataract and presbyopic subjects. Area under the curve equal to 0.781.

9.6 The Times New Roman chart

This is the most common type of reading chart used by optometrists based on the survey results (Chapter 2). It showed a poor repeatability and reliability results for measuring the reading performance metrics in the three groups of subjects. Regarding the reading acuity, despite the variation between young, presbyopic and cataract subjects, all of them were able to read the smallest print size, N6. In addition, the TNR reading speed measurement has a poor ability to discriminate between cataract and presbyopic subjects, as the ROC curve was calculated as 0.650 (Fig 9.6).

Optometrists continue to use many reading charts that have used Times New Roman style with N-notation as a part of the near vision examination despite the fact that it is not developed according to the recommendation of the standardized reading chart and does not use logarithmic progression. Evaluating this type of chart in this thesis was aimed to provide an evidence base with statistical analysis that showed this type of chart is unreliable and should be avoided.

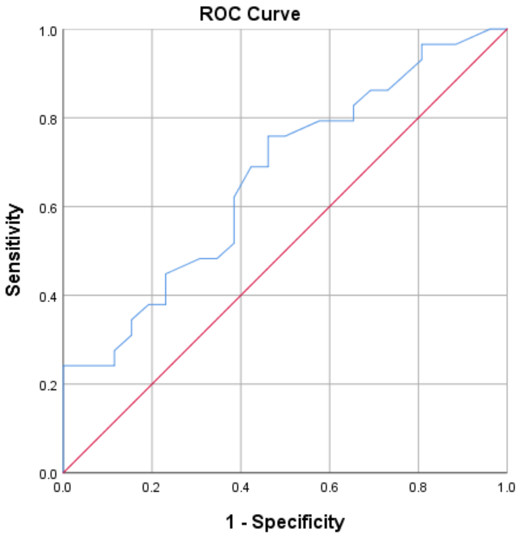


Figure 9.6: The ROC curve showing the ability of the TNR reading speed to discriminate between patients with cataract and presbyopic subjects. Area under the curve equal to 0.650

9.7 Mobile app reading test

This study investigated the reliability and repeatability of two mobile app reading performance tests, Radner and MNread apps, and compared them with traditional paper versions. Overall, the findings indicate that the practitioner should consider that the use of these different two mobile apps is not interchangeable. Different reading outcomes were revealed with different mobile app reading tests due to the variation of the way of administration of the test and the different scoring method used. In addition, the results are not interchangeable with paper-based chart. Using the mobile app test for patients with cataract is not very encouraging since very poor reliability and repeatability of the reading performance metrics have been found.

In normally sighted individuals, the Radner app can be reliable when used to measure the reading performance and it was found to correlate with paper charts better than the MNread app. An implication of the mobile app reading test is that it is ideal due to the compact size and automated method for calculating the reading performance metrics, which may save more time at the examination chair. However, the existing mobile app test needs to be further calibrated and developing more digital reading tests is strongly recommended.

9.8 Best methods for measuring the reading performance in clinical practice

Chapter 8 attempts to provide some critical point for scientific studies or clinical practices that lead to measure the reading performance precisely. First, the reading speed and the reading error results differ between stopwatch recording and audio recording. The stopwatch recording depends on the examiner's reaction time and experience and may produce a source of variability. Both methods are reliable and well accepted, but caution must be applied when comparing the studies. Second, the testing conditions as described as read in armchair or read at desk that are representative of daily practice have no influence on reading speed results. Therefore, determining the appropriate condition based on patient reading behaviour is recommended. Third, clinical trials in ophthalmology often have the distance visual acuity as the main outcome measure. However, it is a poor predictor of reading performance. Advice is given to assess the full reading performance as no single measure of vision has the ability to predict all reading performance metrics. However, the results showed that adding the near CS test (Mars test) to the primary clinical examination had a value in prediction of the RA and CPS in the cataract group of subjects. Finally, uniformity of near visual acuity measures is important, as different near optotypes resulted in different near visual acuity outcomes. The distinction between the "isolated optotype near visual acuity" and "reading acuity" should be adhered to more strictly as no agreement was found between them.

9.9 Standardized reading assessment in presbyopia treatment

In presbyopia research, the use of standardized reading charts that allow evaluation of reading performance before and after presbyopia correction is fundamental. The study described in Chapter 6 is an evaluation of the efficacy of using low concentration pilocarpine drops (0.5%) to improve reading performance for presbyopic subjects by performing a full reading assessment using the Radner reading chart. The results confirmed that the improvement in the near VA measured with ETDRS does not reflect the improvement of the reading performance metrics RA, MRS and CPS. This experiment raises the importance of the use of such standardized reading test charts to evaluate the efficacy of ocular treatment efficiently and accurately.

9.10 Limitations

In the studies described in this thesis the testing method was monocular. This could be considered a limitation since in 'real-life' situations patients are more likely to read binocular. However, from a method point of view the monocular testing could overcome subtle binocular vision problems that may confounded the results near the end of the testing session. Also, to facilitate comparison with other studies as the majority of them used monocular testing. Another limitation, in the design aspects of this study, although a short break of around three minutes was taken before starting the next reading test, the single testing session was long and could produce a source of fatigue effect. Notwithstanding these limitations, the study certainly adds to our understanding that the reading test chart should not be randomly chosen, the importance of use the same type of reading chart for consecutively patients and the unrelated word Bailey-Lovie reading test provides more reliable results than continuous-text reading charts.

9.11 Recommendations for further research work

There are three main recommendations. Firstly, reproducibility analysis (the variability in the repeated measurements when one or more factors are changed, for example, the examiner and the environment) of the reading test charts needs to be further assessed. However, it is important to report the repeatability results of the repeated measurement with a single observer under a laboratory setting to eliminate any source of the variability. This condition may not apply in clinical practice with longitudinal follow-up of the patients with different observers at different settings.

Secondly, the use of a standardized test for the reading performance metrics is of great value in international studies that evaluate the clinical intervention of the diseases affecting the reading ability. Therefore, it is recommended to investigate the reliability and repeatability of the reading test charts in every single language under controlling measurement-related factors that may affect the results.

Thirdly, as the primary concern of the low vision patient is the reading difficulty, direct comparison between the reading test charts in one calibrated study needs to be further assessed. More information on the reliability of the reading test charts would help to establish a greater degree of the chart’s properties.

References

AARÅS, A., HORGEN, G., BJØRSET, H.-H., RO, O. & THORESEN, M. 1998. Musculoskeletal, visual and psychosocial stress in VDU operators before and after multidisciplinary ergonomic interventions. *Applied Ergonomics,* 29**,** 335-354.

ABDELKADER, A. 2015. Improved presbyopic vision with miotics. *Eye & contact lens,* 41**,** 323-327.

AHN, S. J. & LEDGE, G. E. 1995. Psychophysics of reading—XIII. Predictors of magnifier-aided reading speed in low vision. *Vision Research,* 35**,** 1931-1938.

AHN, S. J., LEGGE, G. E. & LUEBKER, A. 1995. Printed cards for measuring low-vision reading speed. *Vision research,* 35**,** 1939-1944.

ALABDULKADER, B. & LEAT, S. J. 2017. Toward developing a standardized Arabic continuous text reading chart. *Journal of optometry,* 10**,** 84-94.

ALEXANDER, M. F., MAGUIRE, M. G., LIETMAN, T. M., SNYDER, J. R., ELMAN, M. J. & FINE, S. L. 1988. Assessment of visual function in patients with age-related macular degeneration and low visual acuity. *Archives of Ophthalmology,* 106**,** 1543-1547.

ALIO, J. L., PLAZA-PUCHE, A. B., PINERO, D. P., AMPARO, F., JIMENEZ, R., RODRIGUEZ-PRATS, J. L., JAVALOY, J. & PONGO, V. 2011. Optical analysis, reading performance, and quality-of-life evaluation after implantation of a diffractive multifocal intraocular lens. *Journal of Cataract and Refractive Surgery,* 37**,** 27-37.

ALIÓ, J. L., RADNER, W., PLAZA-PUCHE, A. B., ORTIZ, D., NEIPP, M. C., QUILES, M. J. & RODRÍGUEZ-MARÍN, J. 2008. Design of short Spanish sentences for measuring reading performance: Radner-Vissum test. *Journal of Cataract & Refractive Surgery,* 34**,** 638-642.

ALTPETER, E. K., MARX, T., NGUYEN, N. X., NAUMANN, A. & TRAUZETTEL-KLOSINSKI, S. 2015. Measurement of reading speed with standardized texts: a comparison of single sentences and paragraphs. *Graefes Arch Clin Exp Ophthalmol,* 253**,** 1369-75.

ARDITI, A. 2005. Improving the design of the letter contrast sensitivity test. *Investigative ophthalmology & visual science,* 46**,** 2225-2229.

ATCHISON, D. A., SCHMID, K. L., EDWARDS, K. P., MULLER, S. M. & ROBOTHAM, J. 2001. The effect of under and over refractive correction on visual performance and spectacle lens acceptance. *Ophthalmic and Physiological Optics,* 21**,** 255-261.

BAILEY, I. L. & LOVIE-KITCHIN, J. E. 2013. Visual acuity testing. From the laboratory to the clinic. *Vision Research,* 90**,** 2-9.

BAILEY, I. L. & LOVIE, J. E. 1976. New design principles for visual acuity letter charts. *American journal of optometry and physiological optics,* 53**,** 740-745.

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.

BALOTA, D. A., POLLATSEK, A. & RAYNER, K. 1985. The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology,* 17**,** 364-390.

BASKARAN, K., MACEDO, A. F., HE, Y., HERNANDEZ-MORENO, L., QUEIRÓS, T., MANSFIELD, J. S. & CALABRÈSE, A. 2019. Scoring reading parameters: An inter-rater reliability study using the MNREAD chart. *PloS one,* 14.

BAUDU, P., PENIN, F. & MOSQUERA, S. A. 2013. Uncorrected binocular performance after biaspheric ablation profile for presbyopic corneal treatment using AMARIS with the PresbyMAX module. *American journal of ophthalmology,* 155**,** 636-647. e1.

BEGIER, E. M., SOCKWELL, D., BRANCH, L. M., DAVIES-COLE, J. O., JONES, L. H., EDWARDS, L., CASANI, J. A. & BLYTHE, D. 2003. The National Capitol Region's Emergency Department syndromic surveillance system: do chief complaint and discharge diagnosis yield different results? *Emerg Infect Dis,* 9**,** 393-6.

BENNETT, A. G. & RABBETTS, R. B. 1998. *Bennett and Rabbetts' Clinical Visual Optics*, Butterworth-Heinemann.

BERGER, E. 2010. The iPad: gadget or medical godsend? *Annals of emergency medicine,* 56**,** A21-A22.

BERGQVIST, U. O. & KNAVE, B. G. 1994. Eye discomfort and work with visual display terminals. *Scandinavian Journal of Work, Environment & Health***,** 27-33.

BLANCHARD, H. E., POLLATSEK, A. & RAYNER, K. 1989. The acquisition of parafoveal word information in reading. *Perception & Psychophysics,* 46**,** 85-94.

BLAND, J. M. & ALTMAN, D. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet,* 327**,** 307-310.

BLAND, J. M. & ALTMAN, D. G. 1996. Measurement error. *Bmj,* 312**,** 1654.

BOUMA, H. 1971. Visual recognition of isolated lower-case letters. *Vision research,* 11**,** 459-474.

BOURNE, R. R. A., FLAXMAN, S. R., BRAITHWAITE, T., CICINELLI, M. V., DAS, A., JONAS, J. B., KEEFFE, J., KEMPEN, J. H., LEASHER, J., LIMBURG, H., NAIDOO, K., PESUDOVS, K., RESNIKOFF, S., SILVESTER, A., STEVENS, G. A., TAHHAN, N., WONG, T. Y., TAYLOR, H. R., BOURNE, R., ACKLAND, P., ARDITI, A., BARKANA, Y., BOZKURT, B., BRAITHWAITE, T., BRON, A., BUDENZ, D., CAI, F., CASSON, R., CHAKRAVARTHY, U., CHOI, J., CICINELLI, M. V., CONGDON, N., DANA, R., DANDONA, R., DANDONA, L., DAS, A., DEKARIS, I., DEL MONTE, M., DEVA, J., DREER, L., ELLWEIN, L., FRAZIER, M., FRICK, K., FRIEDMAN, D., FURTADO, J., GAO, H., GAZZARD, G., GEORGE, R., GICHUHI, S., GONZALEZ, V., HAMMOND, B., HARTNETT, M. E., HE, M., HEJTMANCIK, J., HIRAI, F., HUANG, J., INGRAM, A., JAVITT, J., JONAS, J., JOSLIN, C., KEEFFE, J., KEMPEN, J., KHAIRALLAH, M., KHANNA, R., KIM, J., LAMBROU, G., LANSINGH, V. C., LANZETTA, P., LEASHER, J., LIM, J., LIMBURG, H., MANSOURI, K., MATHEW, A., MORSE, A., MUNOZ, B., MUSCH, D., NAIDOO, K., NANGIA, V., PALAIOU, M., PARODI, M. B., PENA, F. Y., PESUDOVS, K., PETO, T., QUIGLEY, H., RAJU, M., RAMULU, P., RESNIKOFF, S., ROBIN, A., ROSSETTI, L., SAADDINE, J., SANDAR, M. Y. A., SERLE, J., SHEN, T., SHETTY, R., SIEVING, P., SILVA, J. C., SILVESTER, A., SITORUS, R. S., STAMBOLIAN, D., STEVENS, G., et al. 2017. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health,* 5**,** e888-e897.

BROWN, D., DOUGHERTY, P., GILLS, J. P., HUNKELER, J., SANDERS, D. R. & SANDERS, M. L. 2009. Functional reading acuity and performance: comparison of 2 accommodating intraocular lenses. *Journal of Cataract & Refractive Surgery,* 35**,** 1711-1714.

BROWN, J. C., GOLDSTEIN, J. E., CHAN, T. L., MASSOF, R., RAMULU, P. & GROUP, L. V. R. N. S. 2014. Characterizing functional complaints in patients seeking outpatient low-vision services in the United States. *Ophthalmology,* 121**,** 1655-1662. e1.

BRUSSEE, T., VAN DEN BERG, T. J., VAN NISPEN, R. M., DE BOER, I. & VAN RENS, G. H. 2018. Association between contrast sensitivity and reading with macular pathology. *Optometry and Vision Science,* 95**,** 183-192.

BRUSSEE, T., VAN NISPEN, R. M. & VAN RENS, G. H. 2014. Measurement properties of continuous text reading performance tests. *Ophthalmic and Physiological Optics,* 34**,** 636-657.

BRUSSEE, T., VAN NISPEN, R. M. A., KLERKX, E., KNOL, D. L. & VAN RENS, G. 2015. Comparison of reading performance tests concerning difficulty of sentences and paragraphs and their reliability. *Ophthalmic and Physiological Optics,* 35**,** 324-335.

BULLIMORE, M. A. & BAILEY, I. L. 1995. Reading and eye movements in age-related maculopathy. *Optometry and vision science : official publication of the American Academy of Optometry,* 72**,** 125-138.

BURGGRAAFF, M. C., VAN NISPEN, R. M., HOEBEN, F. P., KNOL, D. L. & VAN RENS, G. H. 2012. Randomized controlled trial on the effects of training in the use of closed-circuit television on reading performance. *Investigative ophthalmology & visual science,* 53**,** 2142-2150.

BURGGRAAFF, M. C., VAN NISPEN, R. M., HOEK, S., KNOL, D. L. & VAN RENS, G. H. 2010a. Feasibility of the Radner Reading Charts in low-vision patients. *Graefe's Archive for Clinical and Experimental Ophthalmology,* 248**,** 1631-1637.

BURGGRAAFF, M. C., VAN NISPEN, R. M. A., HOEK, S., KNOL, D. L. & VAN RENS, G. 2010b. Feasibility of the Radner Reading Charts in low-vision patients. *Graefes Archive for Clinical and Experimental Ophthalmology,* 248**,** 1631-1637.

CALABRESE, A., CHEONG, A. M., CHEUNG, S.-H., HE, Y., KWON, M., MANSFIELD, J. S., SUBRAMANIAN, A., YU, D. & LEGGE, G. E. 2016. Baseline MNREAD measures for normally sighted subjects from childhood to old age. *Investigative ophthalmology & visual science,* 57**,** 3836-3843.

CALABRESE, A., GAMAM, S., MANSFIELD, J. S. & LEGGE, G. E. 2014. Implementing the MNREAD reading acuity test on an iPad3. *READING,* 200**,** 300.

CALABRÈSE, A., TO, L., HE, Y., BERKHOLTZ, E., RAFIAN, P. & LEGGE, G. E. 2018. Comparing performance on the MNREAD iPad application with the MNREAD acuity chart. *Journal of vision,* 18**,** 8-8.

CALOSSI, A., BOCCARDO, L., FOSSETTI, A. & RADNER, W. 2014. Design of short Italian sentences to assess near vision performance. *Journal of optometry,* 7**,** 203-209.

CARAMOY, A., FAUSER, S. & KIRCHHOF, B. 2011. Retinal stimuli can be restored after autologous transplant of retinal pigment epithelium and choroid in pigment epithelium tears. *Acta ophthalmologica,* 89**,** e490-e495.

CARVER, R. P. 1976. Word length, prose difficulty, and reading rate. *Journal of Reading Behavior,* 8**,** 193-203.

CASTRO, C., KALLIE, C. & SALOMÃO, S. 2005. Development and validation of the MNREAD reading acuity chart in Portuguese. *Arquivos brasileiros de oftalmologia,* 68**,** 777-783.

CHUNG, S. T., JARVIS, S. H. & CHEUNG, S.-H. 2007. The effect of dioptric blur on reading performance. *Vision research,* 47**,** 1584-1594.

CHYLACK, L. T., JR, WOLFE, J. K., SINGER, D. M., LESKE, M. C., BULLIMORE, M. A., BAILEY, I. L., FRIEND, J., MCCARTHY, D. & WU, S.-Y. 1993. The Lens Opacities Classification System III. *Archives of Ophthalmology,* 111**,** 831-836.

CICCHETTI, D. V. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment,* 6**,** 284.

COLENBRANDER, A. & COMMITTEE, C. O. U. V. F. 1988. Visual acuity measurement standard. *Ital J Ophthalmol,* 2**,** 1-15.

COLENBRANDER, A. & RUNGE, P. 2007. Can Jaeger numbers be standardized. *Investigative Ophthalmology & Visual Science,* 48**,** 3563-3563.

CORNELISSEN, F. W., KOOIJMAN, A. C., VAN SCHOOT, E. A., BOOTSMA, A. & VAN DER WILDT, G.-J. 1994. Optimizing illumination for visually impaired persons; comparing subjective and objective criteria. *Low vision: Research and new developments in rehabilitation***,** 68-77.

DANDONA, R., DANDONA, L., SRINIVAS, M., GIRIDHAR, P., MCCARTY, C. A. & RAO, G. N. 2002. Population-based assessment of refractive error in India: the Andhra Pradesh eye disease study. *Clinical & Experimental Ophthalmology,* 30**,** 84-93.

DAVISON, J. A. & CHYLACK, L. T. 2003. Clinical application of the lens opacities classification system III in the performance of phacoemulsification. *Journal of Cataract & Refractive Surgery,* 29**,** 138-145.

DE WIT, D. W., DIAZ, J. M., MOORE, T. C. & MOORE, J. E. 2012. Refractive lens exchange for a multifocal intraocular lens with a surface-embedded near section in mild to moderate anisometropic amblyopic patients. *Journal of Cataract & Refractive Surgery,* 38**,** 1796-1801.

DELATTRE, P. 1965. *Comparing the phonetic features of English, German, Spanish and French: An interim report*, Julius Groos.
DEXL, A. K., SCHLÖGEL, H., WOLFBAUER, M. & GRABNER, G. 2010. Device for improving quantification of reading acuity and reading speed. *Journal of Refractive Surgery,* 26**,** 682-688.
DOUGHERTY, B. E., FLOM, R. E. & BULLIMORE, M. A. 2005. An evaluation of the Mars letter contrast sensitivity test. *Optometry and Vision Science,* 82**,** 970-975.
EDINGTON, M., SACHDEV, A., MORJARIA, R. & CHONG, V. 2017. STRUCTURAL–FUNCTIONAL CORRELATION IN PATIENTS WITH DIABETIC MACULAR EDEMA. *Retina,* 37**,** 881-885.
ELLIOTT, D., BULLIMORE, M. & BAILEY, I. 1991. Improving the reliability of the Pelli-Robson contrast sensitivity test. *Clinical Vision Sciences,* 6**,** 471-475.
ELLIOTT, D. B., PATEL, B. & WHITAKER, D. 2001. Development of a Reading Speed Test for Potential-Vision Measurements. *Investigative Ophthalmology & Visual Science,* 42**,** 1945-1949.
ELLIOTT, D. B., PATLA, A. E., FURNISS, M. & ADKIN, A. 2000. Improvements in clinical and functional vision and quality of life after second eye cataract surgery. *Optometry and Vision Science,* 77**,** 13-24.
FINE, E. M. & RUBIN, G. S. 1999. The effects of simulated cataract on reading with normal vision and simulated central scotoma. *Vision Research,* 39**,** 4274-4285.
FINGER, R. P., ISSA, P. C., FIMMERS, R., HOLZ, F. G., RUBIN, G. S. & SCHOLL, H. P. 2009. Reading performance is reduced by parafoveal scotomas in patients with macular telangiectasia type 2. *Investigative ophthalmology & visual science,* 50**,** 1366-1370.
FONDA, G. & ANDERSON, M. 1988. Fonda-Anderson reading chart for normal and low vision. *Annals of ophthalmology,* 20**,** 136-139.
FREEMAN, C. E. & EVANS, B. J. W. 2010. Investigation of the causes of non-tolerance to optometric prescriptions for spectacles. *Ophthalmic and Physiological Optics,* 30**,** 1-11.
FRICKE, T. R., TAHHAN, N., RESNIKOFF, S., PAPAS, E., BURNETT, A., HO, S. M., NADUVILATH, T. & NAIDOO, K. S. 2018. Global Prevalence of Presbyopia and Vision Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling. *Ophthalmology,* 125**,** 1492-1499.
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. *Invest Ophthalmol Vis Sci,* 40**,** 858-64.
FUJITA, K., ODA, K., WATANABE, J. & YUZAWA, M. 2008. How normal eyes perform in reading low-contrast texts. *Japanese journal of ophthalmology,* 52**,** 44-47.
FULCHER, G. 1997. Text difficulty and accessibility: reading formulae and expert judgement. *System,* 25**,** 497-513.
GARCÍA-LÁZARO, S., FERRER-BLASCO, T., RADHAKRISHNAN, H., CERVIÑO, A., CHARMAN, W. N. & MONTÉS-MICÓ, R. 2012. Visual function through 4 contact lens–based pinhole systems for presbyopia. *Journal of Cataract & Refractive Surgery,* 38**,** 858-865.
GIACOMELLI, G., VIRGILI, G., GIANSANTI, F., SATO, G., CAPPELLO, E., CRUCIANI, F., VARANO, M. & MENCHINI, U. 2013. Clinical and microperimetric predictors of reading speed in low vision patients: a structural equation modeling approach. *Investigative ophthalmology & visual science,* 54**,** 4403-4408.
GLENI, A., KTISTAKIS, E., TSILIMBARIS, M. K., SIMOS, P., TRAUZETTEL-KLOSINSKI, S. & PLAINIS, S. 2019. Assessing Variability in Reading Performance with the New Greek Standardized Reading Speed Texts (IReST). *Optometry and Vision Science,* 96**,** 761-767.

GREEN, J. 1868. On a new series of test-letters for determining the acuteness of vision. *Transactions of the American Ophthalmological Society,* 1**,** 68.

GRIFFITHS, P. G. & EDDYSHAW, D. 2004. Medically unexplained visual loss in adult patients. *Eye,* 18**,** 917-922.

GRIMM, W., RASSOW, B., WESEMANN, W., SAUR, K. & HILZ, R. 1994. Correlation of optotypes with the Landolt ring--a fresh look at the comparability of optotypes. *Optometry and vision science: official publication of the American Academy of Optometry,* 71**,** 6-13.

GROSVENOR, T. & GROSVENOR, T. P. 2007. *Primary Care Optometry*, Butterworth-Heinemann/Elsevier.

GUPTA, N., NAROO, S. A. & WOLFFSOHN, J. S. 2009a. Visual comparison of multifocal contact lens to monovision. *Optometry and Vision Science,* 86**,** E98-E105.

GUPTA, N., WOLFFSOHN, J. S. & NAROO, S. A. 2009b. Comparison of near visual acuity and reading metrics in presbyopia correction. *Journal of Cataract & Refractive Surgery,* 35**,** 1401-1409.

GUTTERIDGE, I. F. & COLE, B. L. 2000. The prevalence and symptoms of migraine in a consecutive series of patients attending an optometric practice. *Optometry and Vision Science,* 77**,** 402-411.

GWON, A. & WOLDEMUSSIE, E. 2002. Cholinergic agents in the treatment of presbyopia. Google Patents.

HAEGERSTROM-PORTNOY, G., BRABYN, J., SCHNECK, M. E. & JAMPOLSKY, A. 1997. The SKILL Card. An acuity test of reduced luminance and contrast. Smith-Kettlewell Institute Low Luminance. *Investigative ophthalmology & visual science,* 38**,** 207-218.

HAEGERSTROM-PORTNOY, G., SCHNECK, M. E. & BRABYN, J. A. 1999. Seeing into old age: vision function beyond acuity. *Optom Vis Sci,* 76**,** 141-58.

HAHN, G. A., PENKA, D., GEHRLICH, C., MESSIAS, A., WEISMANN, M., HYVÄRINEN, L., LEINONEN, M., FEELY, M., RUBIN, G., DAUXERRE, C., VITAL-DURAND, F., FEATHERSTON, S., DIETZ, K. & TRAUZETTEL-KLOSINSKI, S. 2006. New standardised texts for assessing reading performance in four European languages. *British Journal of Ophthalmology,* 90**,** 480-484.

HAN, Q.-M., CONG, L.-J., YU, C. & LIU, L. 2017. Developing a Logarithmic Chinese Reading Acuity Chart. *Optometry and Vision Science,* 94**,** 714-724.

HASHEMI, N., MOGHADDASI, H., RABIEI, R., ASADI, F. & FARAHI, A. 2018. Eye Health Information Systems in Selected Countries. *Journal of Ophthalmic & Vision Research,* 13**,** 333-338.

HAYMES, S. A., ROBERTS, K. F., CRUESS, A. F., NICOLELA, M. T., LEBLANC, R. P., RAMSEY, M. S., CHAUHAN, B. C. & ARTES, P. H. 2006. The letter contrast sensitivity test: clinical evaluation of a new design. *Investigative ophthalmology & visual science,* 47**,** 2739-2745.

HAZEL, C. A. & ELLIOTT, D. B. 2002. The dependency of logMAR visual acuity measurements on chart design and scoring rule. *Optometry and vision science,* 79**,** 788-792.

HAZEL, C. A., PETRE, K. L., 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 & Visual Science,* 41**,** 1309-1315.

HIRVELÄ, H., KOSKELA, P. & LAATIKAINEN, L. 1995. Visual acuity and contrast sensitivity in the elderly. *Acta Ophthalmologica Scandinavica,* 73**,** 111-115.

HSIEH, S. Y., THOMAS, D. & ROTEM, A. 2005. The organisational response to patient complaints: a case study in Taiwan. *International Journal of Health Care Quality Assurance,* 18**,** 308-320.

İDIL, Ş. A., ÇALIŞKAN, D. & İDİL, N. B. 2011. Development and validation of the Turkish version of the MNREAD visual acuity charts. *Turkish Journal of Medical Sciences,* 41**,** 565-570.

III, F. L. F., FREIDLIN, V., KASSOFF, A., GREEN, S. B. & MILTON, R. C. 1993. Relative letter and position difficulty on visual acuity charts from the Early Treatment Diabetic Retinopathy Study. *American Journal of Ophthalmology,* 116, 735-740.

IRVING, E. L., HARRIS, J. D., MACHAN, C. M., ROBINSON, B. E., HRYNCHAK, P. K., LEAT, S. J. & LILLAKAS, L. 2016. Value of Routine Eye Examinations in Asymptomatic Patients. *Optometry and Vision Science,* 93, 660-666.

JIMMY D. BARTLETT, O. D. D. O. S. S. D. & SIRET D. JAANUS, P. D. L. H. D. 2007. *Clinical Ocular Pharmacology*, Elsevier Health Sciences.

JIN, Y. P., BUYS, Y. M., XIONG, J. & TROPE, G. E. 2013. Government-insured routine eye examinations and prevalence of nonrefractive vision problems among elderly. *Canadian Journal of Ophthalmology-Journal Canadien D Ophtalmologie,* 48, 167-172.

JOSE, R. T. & ATCHERSON, R. 1977. Type-size variability for near-point acuity tests. *American journal of optometry and physiological optics,* 54, 634-638.

KESLER, A., SHEMESH, G., ROTHKOFF, L. & LAZAR, M. 2004. Effect of brimonidine tartrate 0.2% ophthalmic solution on pupil size. *Journal of Cataract & Refractive Surgery,* 30, 1707-1710.

KINGSNORTH, A., DREW, T., GREWAL, B. & WOLFFSOHN, J. S. 2016. Mobile app Aston contrast sensitivity test. *Clinical and Experimental Optometry,* 99, 350-355.

KINGSNORTH, A. & WOLFFSOHN, J. S. 2015. Mobile app reading speed test. *British Journal of Ophthalmology,* 99, 536-539.

KISS, C. G., BARISANI-ASENBAUER, T., MACA, S., RICHTER-MUEKSCH, S. & RADNER, W. 2006. Reading performance of patients with uveitis-associated cystoid macular edema. *American journal of ophthalmology,* 142, 620-624. e1.

KOCH, K. R., MUETHER, P. S., HERMANN, M. M., HOERSTER, R., KIRCHHOF, B. & FAUSER, S. 2012. Subjective perception versus objective outcome after intravitreal ranibizumab for exudative AMD. *Graefe's Archive for Clinical and Experimental Ophthalmology,* 250, 201-209.

KOTTNER, J., AUDIGE, L., BRORSON, S., DONNER, A., GAJEWSKI, B. J., HROBJARTSSON, A., ROBERTS, C., SHOUKRI, M. & STREINER, D. L. 2011. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology,* 64, 96-106.

KRILL, A. E. & NEWELL, F. W. 1964. Effects of pilocarpine on ocular tension dynamics. *American journal of ophthalmology,* 57, 34-41.

LATHAM, K. & WHITAKER, D. 1996. A comparison of word recognition and reading performance in foveal and peripheral vision. *Vision Research,* 36, 2665-2674.

LATINOVIC, R., GULLIFORD, M. & RIDSDALE, L. 2006. Headache and migraine in primary care: consultation, prescription, and referral rates in a large population. *Journal of Neurology, Neurosurgery & Psychiatry,* 77, 385-387.

LAW, F. W. 1952. Reading types. *The British journal of ophthalmology,* 36, 689.

LEAT, S. J. & WOODHOUSE, J. M. 1993. Reading performance with low vision aids: relationship with contrast sensitivity. *Ophthalmic and Physiological Optics,* 13, 9-16.

LEGGE, G. E. 2006. *Psychophysics of Reading in Normal and Low Vision*, CRC Press.

LEGGE, G. E., HOOVEN, T. A., KLITZ, T. S., STEPHEN MANSFIELD, J. & TJAN, B. S. 2002. Mr. Chips 2002: new insights from an ideal-observer model of reading. *Vision Research,* 42, 2219-2234.

LEGGE, G. E., ROSS, J. A., ISENBERG, L. M. & LAMAY, J. M. 1992. Psychophysics of reading. Clinical predictors of low-vision reading speed. *Investigative Ophthalmology & Visual Science,* 33, 677-687.

LEGGE, G. E., ROSS, J. A., LUEBKER, A. & LAMAY, J. M. 1989. Psychophysics of reading. VIII. The Minnesota low-vision reading test. *Optom Vis Sci,* 66, 843-53.

LI, Y.-J., XIRASAGAR, S., PUMKAM, C., KRISHNASWAMY, M. & BENNETT, C. L. 2013. Vision Insurance, Eye Care Visits, and Vision Impairment Among Working-Age Adults in the United StatesVision Insurance, Eye Visits, Vision Impairment. *JAMA Ophthalmology,* 131, 499-506.

LIVINGSTONE, I., TARBERT, C., GIARDINI, M., BASTAWROUS, A., MIDDLETON, D. & HAMILTON, R. 2016. Photometric compliance of tablet screens and retro-illuminated acuity charts as visual acuity measurement devices. *PloS one,* 11.

LOTT, L. A., SCHNECK, M. E., HAEGERSTROM-PORTNOY, G., BRABYN, J. A., GILDENGORIN, G. L. & WEST, C. G. 2001a. Reading performance in older adults with good acuity. *Optom Vis Sci,* 78**,** 316-24.

LOTT, L. A., SCHNECK, M. E., HAEGERSTRÖM-PORTNOY, G., BRABYN, J. A., GILDENGORIN, G. L. & WEST, C. G. 2001b. Reading performance in older adults with good acuity. *Optometry and Vision Science,* 78**,** 316-324.

LOVIE-KITCHIN, J. E. & BROWN, B. 2000. Repeatability and intercorrelations of standard vision tests as a function of age. *Optometry and vision science,* 77**,** 412-420.

MAAIJWEE, K., MULDER, P., RADNER, W. & VAN MEURS, J. C. 2008a. Reliability testing of the Dutch version of the Radner Reading Charts. *Optometry and Vision Science,* 85**,** 353-358.

MAAIJWEE, K., MULDER, P., RADNER, W. & VAN MEURS, J. C. 2008b. Reliability testing of the Dutch version of the Radner Reading Charts. *Optom Vis Sci,* 85**,** 353-8.

MAAIJWEE, K. J., MEULENDIJKS, C. F., RADNER, W., VAN MEURS, J. C. & HOYNG, C. B. 2007. [The Dutch version of the Radner Reading Chart for assessing vision function]. *Nederlands tijdschrift voor geneeskunde,* 151**,** 2494-2497.

MACKEBEN, M., NAIR, U. K. W., WALKER, L. L. & FLETCHER, D. C. 2015. Random Word Recognition Chart Helps Scotoma Assessment in Low Vision. *Optometry and Vision Science,* 92**,** 421-428.

MANGIONE, C. M., BERRY, S., SPRITZER, K., JANZ, N. K., KLEIN, R., OWSLEY, C. & LEE, P. P. 1998. Identifying the content area for the 51-item National Eye Institute Visual Function Questionnaire: results from focus groups with visually impaired persons. *Arch Ophthalmol,* 116**,** 227-33.

MANSFIELD, J., AHN, S., LEGGE, G. & LUEBKER, A. 1993. A new reading-acuity chart for normal and low vision. *Ophthalmic and Visual Optics/Noninvasive Assessment of the Visual System Technical Digest,* 3**,** 232-235.

MARUMOTO, T., SOTOYAMA, M., VILLANUEVA, M. B. G., JONAI, H., YAMADA, H., KANAI, A. & SAITO, S. 1998. Relationship between posture and myopia among students. *Myopia updates.* Springer.

MATAFTSI, A., BOURTOULAMAIOU, A., HAIDICH, A.-B., ANTONIADIS, A., KILINTZIS, V., TSINOPOULOS, I. T. & DIMITRAKOS, S. 2013. Development and validation of the Greek version of the MNREAD acuity chart. *Clinical and Experimental Optometry,* 96**,** 25-31.

MCALINDEN, C., KHADKA, J. & PESUDOVS, K. 2011. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic and Physiological Optics,* 31**,** 330-338.

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.

MCCULLOUGH, S. J., O'DONOGHUE, L. & SAUNDERS, K. J. 2016. Six year refractive change among white children and young adults: evidence for significant increase in myopia among white UK children. *PloS one,* 11**,** e0146332.

MCMONNIES, C. W. 2001. Chart memory and visual acuity measurement. *Clinical and Experimental Optometry,* 84**,** 26-34.

MONES, J. & RUBIN, G. 2005. Contrast sensitivity as an outcome measure in patients with subfoveal choroidal neovascularisation due to age-related macular degeneration. *Eye,* 19**,** 1142-1150.

MORTON, J. 1964. The Effects of Context upon Speed of Reading, Eye Movements and Eye-voice Span. *Quarterly Journal of Experimental Psychology,* 16**,** 340-354.

MOWAFI, H., DWORKIS, D., BISANZO, M., HANSOTI, B., SEIDENBERG, P., OBERMEYER, Z., HAUSWALD, M. & REYNOLDS, T. A. 2013. Making Recording and Analysis of Chief Complaint a Priority for Global Emergency Care Research in Low-income Countries. *Academic Emergency Medicine,* 20**,** 1241-1245.

MUNCH, I. C., JØRGENSEN, A.-H. R. & RADNER, W. 2016. The Danish version of the Radner Reading Chart: design and empirical testing of sentence optotypes in subjects of varying educational background. *Acta Ophthalmologica,* 94**,** 182-186.

NAIR, U. K., MACKEBEN, M., SCHUCHARD, R. A., WATSON, G., FU, A. & JR., D. C. F. 2006. Random Word Reading Test Like MNread Shows More Frequent Errors Than Continuous Text. *Investigative Ophthalmology & Visual Science,* 47**,** 3481-3481.

NAKAZAWA, T., OKUBO, Y., SUWAZONO, Y., KOBAYASHI, E., KOMINE, S., KATO, N. & NOGAWA, K. 2002. Association between duration of daily VDT use and subjective symptoms. *American Journal of Industrial Medicine,* 42**,** 421-426.

NANDY, A. S. & TJAN, B. S. 2012. Saccade-confounded image statistics explain visual crowding. *Nature neuroscience,* 15**,** 463-469.

NAROO, S. A. & BILKHU, P. S. 2016. Clinical utility of the KAMRA corneal inlay. *Clinical ophthalmology (Auckland, NZ),* 10**,** 913.

NAS-NRC 1980. Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of Working Group 39. Committee on Vision. Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, DC. *Advances in Ophthalmology,* 41**,** 103-148.

NEUGEBAUER, A., FRICKE, J. & RÜSSMANN, W. 1992. Asthenopia: frequency and objective findings. *German journal of ophthalmology,* 1**,** 122-124.

NGUYEN, N. X., WEISMANN, M. & TRAUZETTEL-KLOSINSKI, S. 2009. Improvement of reading speed after providing of low vision aids in patients with age-related macular degeneration. *Acta ophthalmologica,* 87**,** 849-853.

NOLAN, G. M. 2001. Methods for treating various eye disorders. Google Patents.

PATEL, I., MUNOZ, B., BURKE, A. G., KAYONGOYA, A., MCHIWA, W., SCHWARZWALDER, A. W. & WEST, S. K. 2006. Impact of Presbyopia on Quality of Life in a Rural African Setting. *Ophthalmology,* 113**,** 728-734.

PATEL, I. & WEST, S. K. 2007. Presbyopia: prevalence, impact, and interventions. *Community eye health,* 20**,** 40-41.

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. *Invest Ophthalmol Vis Sci,* 52**,** 3854-9.

PATEL, P. J., CHEN, F. K., RUBIN, G. S. & TUFAIL, A. 2008. Intersession repeatability of visual acuity scores in age-related macular degeneration. *Investigative ophthalmology & visual science,* 49**,** 4347-4352.

PEABODY, J. W., LUCK, J., GLASSMAN, P., JAIN, S., HANSEN, J., SPELL, M. & LEE, M. 2004. Measuring the Quality of Physician Practice by Using Clinical Vignettes: A Prospective Validation Study. *Annals of Internal Medicine,* 141**,** 771-780.

PELLI, D. & ROBSON, J. The design of a new letter chart for measuring contrast sensitivity. Clinical Vision Sciences, 1988. Citeseer.

PELLI, D. G. & BEX, P. 2013. Measuring contrast sensitivity. *Vision research,* 90**,** 10-14.

PERRUCCIO, A. V., BADLEY, E. M. & TROPE, G. E. 2007. Self-reported glaucoma in Canada: findings from population-based surveys, 1994–2003. *Canadian Journal of Ophthalmology,* 42**,** 219-226.

PULA, J. 2012. Functional vision loss. *Current Opinion in Ophthalmology,* 23**,** 460-465.

RADNER, W. 2017. Reading charts in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology,* 255**,** 1465-1482.

RADNER, W. & DIENDORFER, G. 2014. English sentence optotypes for measuring reading acuity and speed—the English version of the Radner Reading Charts. *Graefe's Archive for Clinical and Experimental Ophthalmology,* 252**,** 1297-1303.

RADNER, W., DIENDORFER, G., KAINRATH, B. & KOLLMITZER, C. 2017. The accuracy of reading speed measurement by stopwatch versus measurement with an automated computer program (rad-rd©). *Acta ophthalmologica,* 95**,** 211-216.
RADNER, W., OBERMAYER, W., RICHTER-MUEKSCH, S., WILLINGER, U., VELIKAY-PAREL, M. & EISENWORT, B. 2002. The validity and reliability of short German sentences for measuring reading speed. *Graefes Archive for Clinical and Experimental Ophthalmology,* 240**,** 461-467.
RADNER, W., WILLINGER, U., OBERMAYER, W., MUDRICH, C., VELIKAY-PAREL, M. & EISENWORT, B. 1998. A new reading chart for simultaneous determination of reading vision and reading speed. *Klinische Monatsblatter fur Augenheilkunde,* 213**,** 174-181.
RAMULU, P. Y., SWENOR, B. K., JEFFERYS, J. L. & RUBIN, G. S. 2013. Description and Validation of a Test to Evaluate Sustained Silent Reading. *Investigative Ophthalmology & Visual Science,* 54**,** 673-680.
RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin,* 124**,** 372.
RAYNER, K., BINDER, K. S., ASHBY, J. & POLLATSEK, A. 2001. Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research,* 41**,** 943-954.
RAYNER, K., SERENO, S. C. & RANEY, G. E. 1996. Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance,* 22**,** 1188.
RAYNER, K., WELL, A. D., POLLATSEK, A. & BERTERA, J. H. 1982. The availability of useful information to the right of fixation in reading. *Perception & Psychophysics,* 31**,** 537-550.
RESNIKOFF, S., PASCOLINIA, D., MARIOTT, S. P. & POKHAREL, G. P. 2008. Global magnitude of visual impairment caused by uncorrected refractive errors in 2004. *Bulletin of the World Health Organization,* 86**,** 63-70.
RICE, M. L., BIRCH, E. E. & HOLMES, J. M. 2005. An abbreviated reading speed test. *Optometry and vision science,* 82**,** 128-133.
RICHTER-MUEKSCH, S., KAMINSKI, S., KUCHAR, A., STIFTER, E., VELIKAY-PAREL, M. & RADNER, W. 2005. Influence of laser in situ keratomileusis and laser epithelial keratectomy on patients' reading performance. *Journal of Cataract & Refractive Surgery,* 31**,** 1544-1548.
RICHTER-MUEKSCH, S., WEGHAUPT, H., SKORPIK, C., VELIKAY-PAREL, M. & RADNER, W. 2002. Reading performance with a refractive multifocal and a diffractive bifocal intraocular lens. *Journal of Cataract & Refractive Surgery,* 28**,** 1957-1963.
ROBAEI, D., KIFLEY, A., ROSE, K. A. & MITCHELL, P. 2006. Refractive error and patterns of spectacle use in 12-year-old Australian children. *Ophthalmology,* 113**,** 1567-1573.
RUBIN, G. S. 2013. Measuring reading performance. *Vision Research,* 90**,** 43-51.
RUBIN, G. S., WEST, S. K., MUÑOZ, B., BANDEEN-ROCHE, K., ZEGER, S., SCHEIN, O. & FRIED, L. P. 1997. A comprehensive assessment of visual impairment in a population of older Americans. The SEE Study. Salisbury Eye Evaluation Project. *Investigative Ophthalmology & Visual Science,* 38**,** 557-568.
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.
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**,** 166-177.
SCHEIMAN, M. & WICK, B. 2008. *Clinical Management of Binocular Vision: Heterophoric, Accommodative, and Eye Movement Disorders*, Wolters Kluwer Health/Lippincott Williams & Wilkins.
SCHLEYER, T. K. L. & FORREST, J. L. 2000. Methods for the Design and Administration of Web-based Surveys. *Journal of the American Medical Informatics Association,* 7**,** 416-425.

SHAH, R., EDGAR, D. F., RABBETTS, R., BLAKENEY, S. L., CHARLESWORTH, P., HARLE, D. E. & EVANS, B. J. W. 2008. The content of optometric eye examinations for a young myope with headaches. *Ophthalmic and Physiological Optics,* 28**,** 404-421.

SHEEDY, J. E., BAILEY, I. L. & RAASCH, T. W. 1984. Visual acuity and chart luminance. *Optometry and Vision Science,* 61**,** 595-600.

SHEEDY, J. E., SUBBARAM, M. V., ZIMMERMAN, A. B. & HAYES, J. R. 2005. Text legibility and the letter superiority effect. *Human factors,* 47**,** 797-815.

SHICKLE, D. & GRIFFIN, M. 2014. Why don't older adults in England go to have their eyes examined? *Ophthalmic and Physiological Optics,* 34**,** 38-45.

SHICKLE, D., TODKILL, D., CHISHOLM, C., RUGHANI, S., GRIFFIN, M., CASSELS-BROWN, A., MAY, H., SLADE, S. V. & DAVEY, C. J. 2015. Addressing inequalities in eye health with subsidies and increased fees for General Ophthalmic Services in socio-economically deprived communities: a sensitivity analysis. *Public Health,* 129**,** 131-137.

SHINGLETON, B. J. & O'DONOGHUE, M. W. 2000. Primary care: Blurred vision. *New England Journal of Medicine,* 343**,** 556-562.

SHNEOR, E. & HOCHSTEIN, S. 2008. Eye dominance effects in conjunction search. *Vision Research,* 48**,** 1592-1602.

SLOAN, L. L. 1959. New test charts for the measurement of visual acuity at far and near distances. *American journal of ophthalmology,* 48**,** 807-813.

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.

SNELLEN, H. 1862. *Letterproeven, tot bepaling der gezigtsscherpte*, J. Greven.

STANGLER-ZUSCHROTT, E. 1990. [Decreased reading speed and rapid fatigue as signs of disordered visual function]. *Klinische Monatsblatter fur Augenheilkunde,* 196**,** 150-157.

STARR, M. S. & RAYNER, K. 2001. Eye movements during reading: Some current controversies. *Trends in Cognitive Sciences,* 5**,** 156-163.

STIFTER, E., BURGGASSER, G., HIRMANN, E., THALER, A. & RADNER, W. 2005a. Evaluating reading acuity and speed in children with microstrabismic amblyopia using a standardized reading chart system. *Graefe's Archive for Clinical and Experimental Ophthalmology,* 243**,** 1228-1235.

STIFTER, E., KÖNIG, F., LANG, T., BAUER, P., RICHTER-MÜKSCH, S., VELIKAY-PAREL, M. & RADNER, W. 2004a. Reliability of a standardized reading chart system: variance component analysis, test-retest and inter-chart reliability. *Graefe's Archive for Clinical and Experimental Ophthalmology,* 242**,** 31-39.

STIFTER, E., SACU, S., WEGHAUPT, H., KÖNIG, F., RICHTER-MÜKSCH, S., THALER, A., VELIKAY-PAREL, M. & RADNER, W. 2004b. Reading performance depending on the type of cataract and its predictability on the visual outcome. *Journal of Cataract & Refractive Surgery,* 30**,** 1259-1267.

STIFTER, E., WEGHAUPT, H., BENESCH, T., THALER, A. & RADNER, W. 2005b. Discriminative power of reading tests to differentiate visual impairment caused by cataract and age-related macular degeneration. *Journal of Cataract & Refractive Surgery,* 31**,** 2111-2119.

STIFTER, E., WEGHAUPT, H., BENESCH, T., THALER, A. & RADNER, W. 2005c. Discriminative power of reading tests to differentiate visual impairment caused by cataract and age-related macular degeneration. *Journal of Cataract and Refractive Surgery,* 31**,** 2111-2119.

STRASBURGER, H., WÜSTENBERG, T. & JÄNCKE, L. 2002. Calibrated LCD/TFT stimulus presentation for visual psychophysics in fMRI. *Journal of neuroscience methods,* 121**,** 103-110.

SUBRAMANIAN, A. & PARDHAN, S. 2006. The repeatability of MNREAD acuity charts and variability at different test distances. *Optometry and vision science,* 83**,** 572-576.

SUBRAMANIAN, A. & PARDHAN, S. 2009. Repeatability of reading ability indices in subjects with impaired vision. *Invest Ophthalmol Vis Sci,* 50**,** 3643-7.

SUNNESS, J. S. & APPLEGATE, C. A. 2005. Long-term follow-up of fixation patterns in eyes with central scotomas from geographic atrophy that is associated with age-related macular degeneration. *American journal of ophthalmology,* 140**,** 1085-1093.

TAVAKOL, M. & DENNICK, R. 2011. Making sense of Cronbach's alpha. *International journal of medical education,* 2**,** 53-55.

TAYLOR, H. R., VU, H. T. V., MCCARTY, C. A. & KEEFFE, J. E. 2004. The Need for Routine Eye Examinations. *Investigative Ophthalmology & Visual Science,* 45**,** 2539-2542.

TEJERIA, L., HARPER, R., ARTES, P. & DICKINSON, C. 2002. Face recognition in age related macular degeneration: perceived disability, measured disability, and performance with a bioptic device. *British Journal of Ophthalmology,* 86**,** 1019-1026.

THAYAPARAN, K., CROSSLAND, M. D. & RUBIN, G. S. 2007. Clinical assessment of two new contrast sensitivity charts. *British Journal of Ophthalmology,* 91**,** 749-752.

TRAUZETTEL-KLOSINSKI, S., DIETZ, K. & GROUP, T. I. S. 2012. Standardized Assessment of Reading Performance: The New International Reading Speed Texts IReST. *Investigative Ophthalmology & Visual Science,* 53**,** 5452-5461.

VILELA, M. A. P., PELLANDA, L. C., FASSA, A. G. & CASTAGNO, V. D. 2015. Prevalence of asthenopia in children: a systematic review with meta-analysis. *Jornal de Pediatria,* 91**,** 320-325.

VIRGILI, G., CORDARO, C., BIGONI, A., CROVATO, S., CECCHINI, P. & MENCHINI, U. 2004a. Reading Acuity in Children: Evaluation and Reliability Using MNREAD Charts. *Investigative Ophthalmology & Visual Science,* 45**,** 3349-3354.

VIRGILI, G., PIERROTTET, C., PARMEGGIANI, F., PENNINO, M., GIACOMELLI, G., STEINDLER, P., MENCHINI, U. & ORZALESI, N. 2004b. Reading performance in patients with retinitis pigmentosa: a study using the MNREAD charts. *Investigative ophthalmology & visual science,* 45**,** 3418-3424.

WEIH, L. M., HASSELL, J. B. & KEEFFE, J. 2002. Assessment of the Impact of Vision Impairment. *Investigative Ophthalmology & Visual Science,* 43**,** 927-935.

WENSOR, M. D., MCCARTY, C. A., STANISLAVSKY, Y. L., LIVINGSTON, P. M. & TAYLOR, H. R. 1998. The prevalence of glaucoma in the melbourne visual impairment project. *Ophthalmology,* 105**,** 733-739.

WEST, S. K., MUNOZ, B., RUBIN, G. S., SCHEIN, O. D., BANDEEN-ROCHE, K., ZEGER, S., GERMAN, S. & FRIED, L. P. 1997. Function and visual impairment in a population-based study of older adults. The SEE project. Salisbury Eye Evaluation. *Invest Ophthalmol Vis Sci,* 38**,** 72-82.

WILLIAMS, M. A., MOUTRAY, T. N. & JACKSON, A. J. 2008. Uniformity of visual acuity measures in published studies. *Investigative ophthalmology & visual science,* 49**,** 4321-4327.

WITTENBORN, J. S., ZHANG, X., FEAGAN, C. W., CROUSE, W. L., SHRESTHA, S., KEMPER, A. R., HOERGER, T. J. & SAADDINE, J. B. 2013. The Economic Burden of Vision Loss and Eye Disorders among the United States Population Younger than 40 Years. *Ophthalmology,* 120**,** 1728-1735.

WITTICH, W., OVERBURY, O., KAPUSTA, M. A. & WATANABE, D. H. 2006. Differences between recognition and resolution acuity in patients undergoing macular hole surgery. *Investigative ophthalmology & visual science,* 47**,** 3690-3694.

XIONG, Y.-Z., CALABRESE, A., CHEONG, A. M. & LEGGE, G. E. 2018. Reading acuity as a predictor of low-vision reading performance. *Investigative Ophthalmology & Visual Science,* 59**,** 4798-4803.

XU, R. & BRADLEY, A. 2015. IUR ead: a new computer-based reading test. *Ophthalmic and Physiological Optics,* 35**,** 500-513.

YILMAZ, Ö. F., BAYRAKTAR, S., AGCA, A., YILMAZ, B., MCDONALD, M. B. & VAN DE POL, C. 2008. Intracorneal inlay for the surgical correction of presbyopia. *Journal of Cataract & Refractive Surgery,* 34**,** 1921-1927.

ZACZEK, A. & ZETTERSTRÖM, C. 1998. The effect of phenylephrine and pilocarpine on pupil size and aqueous flare intensity in patients with diabetes mellitus. *Acta Ophthalmologica Scandinavica,* 76**,** 413-416.
ZAKI, R., BULGIBA, A., ISMAIL, R. & ISMAIL, N. A. 2012. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PloS one,* 7.
ZERI, F., CALCATELLI, P., FUNARO, E., MARTELLI, M. & NAROO, S. A. 2018. How accurate is an LCD screen version of the Pelli–Robson test? *International ophthalmology,* 38**,** 1473-1484.
ZHAO, J., PAN, X., SUI, R., MUNOZ, S. R., SPERDUTO, R. D. & ELLWEIN, L. B. 2000. Refractive error study in children: results from Shunyi District, China. *American journal of ophthalmology,* 129**,** 427-435.

Appendices

Appendix 1- Types of near vision charts used in optometric practice questionnaire and consent form (Chapter 2)

Aston University
Life & Health Sciences

Statement of Informed Consent Form: Participant Copy

Title of Project: An evaluation of the reading performance and visual acuity tests
Research Venue: Vision Sciences Building, Aston University, B4 7ET
Name of Investigator(s): Dr Shehzad Naroo and Miss Mashaaer Baashen

Dear Practitioner,
My name is Mashaaer Baashen. As a part of my PhD project at Aston University in Birmingham I am carrying out a survey about the types of near vision testing charts and techniques used in optometric practice and would appreciate a few minutes of your time in completing this brief questionnaire. The results will be treated in confidence and we only ask for your name to ensure that we do not use two sets of results for the same person.

Please initial box

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my interaction with the university being affected. ☐
3. I agree to take part in the above study. ☐

________________	________	____________
Name of Research Participant	Date	Signature
Name of Person taking Consent	Date	Signature

Version 1 1 copy for research participant; 1 copy for Aston University

1) Initials of practitioner: ______________________

2) Type of Practice Please tick

Hospital	
Laser Clinic	
Single independent	
Chain (<5 stores)	
Chain (5 to 30 stores)	
Franchise of large multiple	
Large multiple (over 30 stores)	

3) Location of your primary workplace or clinical practice

____

4) When did you qualify?

________ Years ________ Months

5) On average what percentage (%) of the patients do you see each month in the following category?

Pre-school children (under 5) ______

School age children (5-16 years) ______

Pre-presbyopic adults (16-44 years) ______

Presbyopic age 45- 60 years ______

Presbyopic over 60 years ______

6) What is the most common reason(s) for the patient for visiting the optometry clinic?

- ☐ Family history of ocular diseases or refractive errors
- ☐ Routine eye examination
- ☐ Pathology
- ☐ Near difficulties
- ☐ Distance difficulties
- ☐ Spectacles
- ☐ Headache
- ☐ Contact lenses
- ☐ Others ____

7) What style of the distance test chart do you use?

- ☐ Snellen Chart
- ☐ LogMAR Chart
- ☐ Landolt 'C' Chart
- ☐ Tumbling 'E' Chart
- ☐ Others ______________________________________

8) What type of the reading test charts do you routinely use?
Please write full name of the chart

1. ____
2. ____
3. ____

9) What type of near reading test charts are available in your clinic?
Please write full name of chart

1. ____
2. ____
3. ____

END OF QUESTIONNAIRE

Appendix 2 - Bland-Altman agreement plots (Chapter 3)

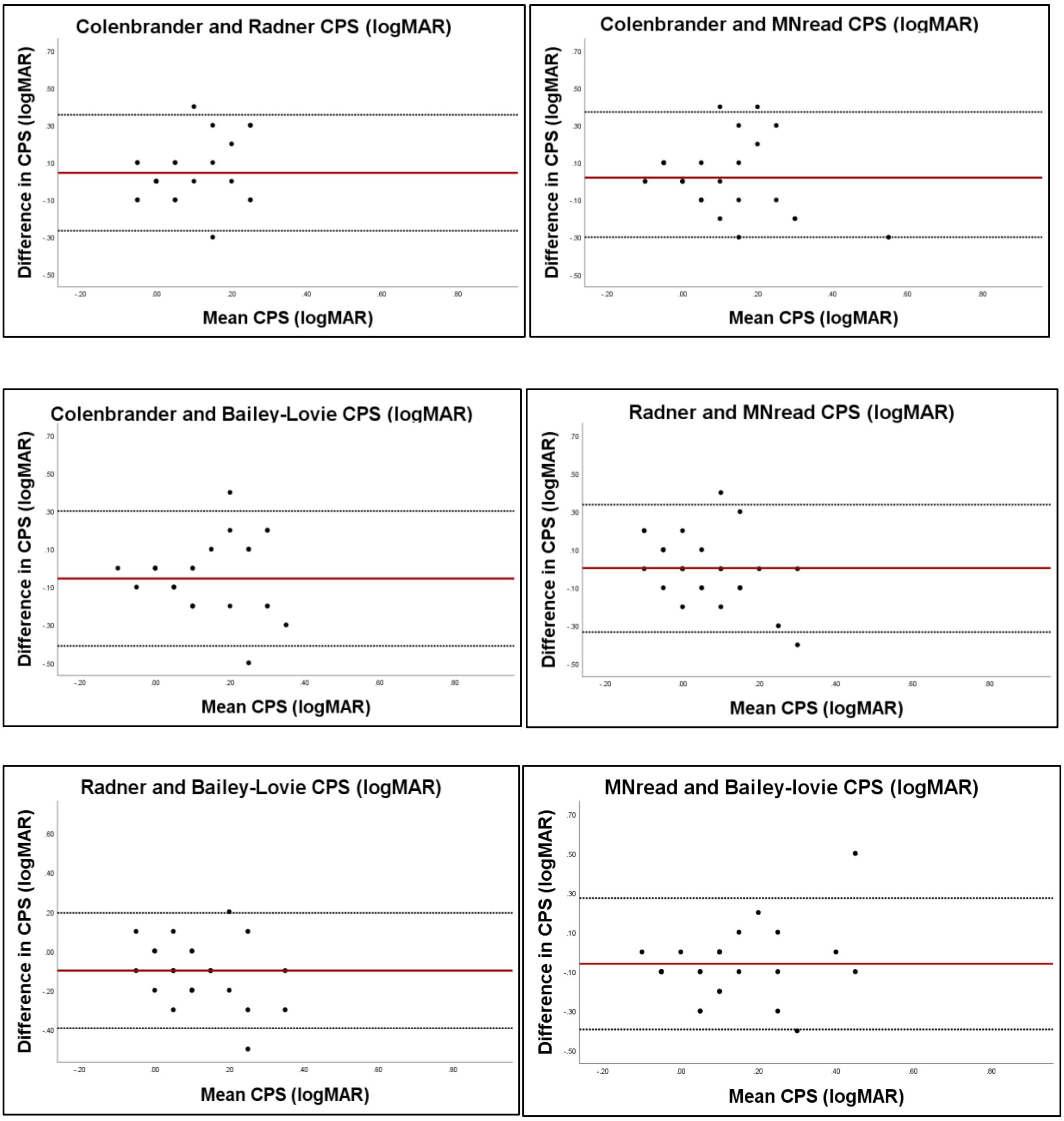


Figure A 2.1: Bland-Altman agreement plots of the CPS between the charts. The differences of the mean between the charts is plotted against the mean CPS of both charts.

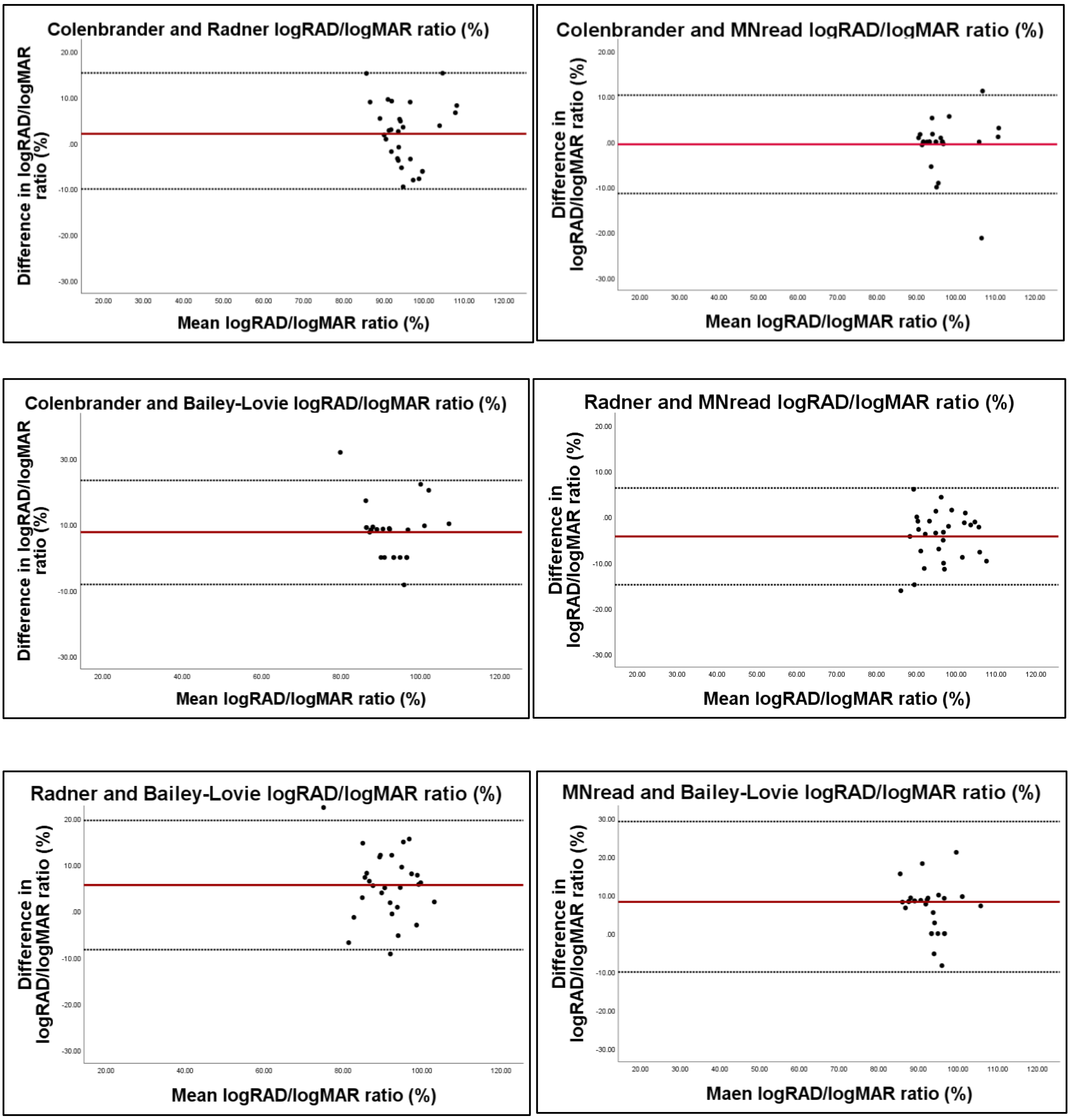


Figure A 2.2: Bland-Altman agreement plots of the logRAD/LogMAR ratio between the charts. The differences of the mean between the charts is plotted against the mean logRAD/logMAR ratio of both charts

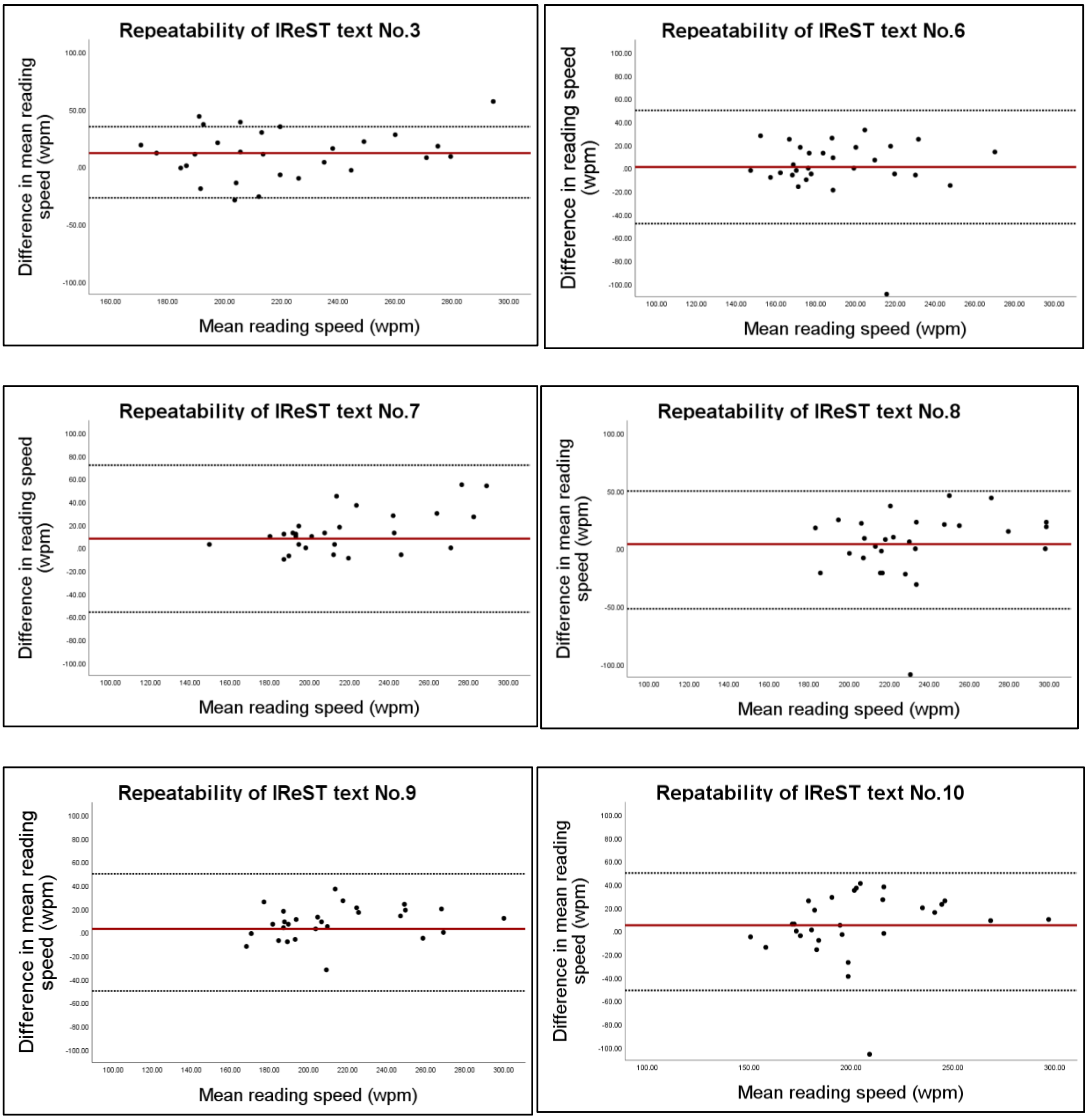


Figure A 2.3: Bland-Altman agreement shows a high agreement between the two testing sessions in IReST text No. 3, 6, 8, 9 and 10, but lower agreement found in text number 7

Appendix 3 - Ethical approval

Aston University

Aston University
Aston Triangle
Birmingham
B4 7ET
0121 204 3000

Date: 13/10/2016

[redacted]
Life and Health Sciences

Study title:	Comparison of different charts and methods for measuring distance and near visual performance in presbyopic and non-presbyopic people
REC REF:	Ethics application #876

Confirmation of Ethical Opinion

On behalf of the Committee, I am pleased to confirm a favourable opinion for the amendment to this research.

Documents approved

Document	*Version*	*Date*
Participant information sheet (VA study PIS v5)	5	12th October 2016
Consent form (VA study consent v6)	6	12th October 2016
Email with amendment request and further explanations/justifications		12th October 2016

With the Committee's best wishes for the success of this project.
Yours sincerely

[redacted]

[redacted]
Chair of the University Research Ethics Committee

Aston University

Personal Identification Number for this study: ____________

CONSENT FORM

Title of Project: Comparison of different charts and methods for measuring distance and near visual performance in presbyopic and non-presbyopic people

Research Venue: Ophthalmic Research Group, Aston University, B4 7ET

Aston University Investigators: Shehzad Naroo MSc, PhD

Fabrizio Zeri, PhD

Mashaaer Baashen BSc (Hons), MSc

Please initial each box

By adding my initials to each box I confirm that:

1. I have read and understand the patient information sheet (version 5) for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary; the study tests are not part of any clinical treatment and do not negate the need for regular eye examinations ☐
3. I understand that I am free to withdraw at any time, without giving any reason and without my interaction with the university being affected ☐
4. I agree to audio recording of my reading ability during the above study ☐
5. I agree to take part in the above study ☐

______________________	______________	__________________
Name of Research Participant	Date	Signature
______________________	______________	__________________
Name of Person taking Consent	Date	Signature

1 copy for research participant; 1 copy for Aston University

Appendix 4 - Ethical approval for pilocarpine study (Chapter 6)

Aston University
Aston Triangle
Birmingham
B4 7ET
0121 204 3000

Date: 11/04/17

Life and Health Sciences

Dear [redacted]

Study title:	An investigation of change in depth of vision with constriction of the pupil
REC REF:	Ethics application -

Confirmation of Ethical Opinion

On behalf of the Committee, I am pleased to confirm a favourable opinion for the above research based on the basis described in the application form, protocol and supporting documentation listed below.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Participant information sheet	1	11/03/17
Study Protocol	1	11/03/17
Study Advertisement	1	28/03/17
Participant Consent Form	1	11/03/17
GP Letter	1	11/03/17

With the Committee's best wishes for the success of this project.
Yours sincerely

[redacted]

Chair of the University Research Ethics Committee

Aston University

An Investigation of Change in Depth of Vision with Constriction of the Pupil

CONSENT FORM

Participant Identification Number: ______

Please initial box

1. I confirm that I have read and understand the information sheet dated 11/03/17 (version 1.0) for the above study. I have had the opportunity to consider the information provided, ask questions and have had these answered to my satisfaction.
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason and my legal rights being affected.
3. I understand that my data may be reviewed by authorised individuals from Aston University responsible for ensuring the quality of the research.
4. I agree to my General Practitioner (GP) being informed of my participation in the study.
5. I agree to take part in the above study.

Name of Participant	Date	Signature
Name of Researcher	Date	Signature

1 copy for the participant, 1 for researcher site file

DOF Study – Consent Form
Version 1 11th March 2017

Page 1 of 1

Appendix 5 - Near Activity Visual Questionnaire (NAVQ)

THE NEAR ACTIVITY VISUAL QUESTIONNAIRE (NAVQ)

Demographic Information

Name: ______________________________ DOB: ___ / ___ / 19___ Gender: Male/Female Date: ___ / ___ / 20___

Type of Correction: (Please Tick ONE Option ONLY)

Spectacles:	❑ Single Vision	❑ Bifocals	❑ Varifocals
Contact Lenses:	❑ Distance CL + Reading Specs	❑ Monovision	❑ Multifocals
IOL:	❑ Single Vision	❑ Multifocal	❑ Accommodating

INSTRUCTIONS – Please read carefully

THIS QUESTIONNAIRE ASKS YOU TO RATE HOW MUCH **DIFFICULTY** YOU EXPERIENCE IN DOING VARIOUS NEAR-VISION RELATED ACTIVITIES.

Please answer ALL questions for **IF/WHEN YOU DO THE DESCRIBED ACTIVITY WITHOUT EXTRA READING OR MAGNIFYING AIDS** (therefore ONLY with your normal correction). Simply CIRCLE THE NUMBER THAT CORRESPONDS TO THE LABEL OF YOUR RESPONSE, as shown in the example:

EXAMPLE:	N/A *or* Stopped for Non-visual Reasons	A Little Difficulty	Moderate Difficulty	Stopped for Visual Reasons
How much difficulty do you have reading road signs when driving?	0	1	2	3

If you do not do the described activity or you have stopped for reasons that are not related to your vision, then please circle the *'N/A'* option.

How much difficulty do you have:	**Not Applicable *Or* Stopped for Non-visual Reasons *Or* No Difficulty**	**A Little Difficulty**	**Moderate Difficulty**	**Extreme Difficulty *Or* Stopped for Visual Reasons**
1. Reading small print, e.g. newspaper articles, books, magazine articles, menus at a restaurant, telephone directories, etc.?	0	1	2	3
2. Reading labels / instructions / ingredients / prices on, e.g. medicine bottles, food packaging, etc.?	0	1	2	3
3. Reading your post / mail, e.g. electric bills, greetings cards, bank statements, letters from friends and family, etc.?	0	1	2	3
4. Seeing the screen & keyboard when using a computer?	0	1	2	3
5. Seeing the display & keypad on a mobile or fixed telephone, or calculator?	0	1	2	3
6. Seeing objects close to you to engage in your hobbies, e.g. playing games such as cards, bingo and dominoes, gardening, seeing photographs and pictures, etc.?	0	1	2	3
7. Doing fine handwork, e.g. threading a needle, sewing, knitting, crochet, etc.?	0	1	2	3
8. Seeing objects near to you in poor or dim light?	0	1	2	3
9. Seeing objects near to you if there is glare from lights and shiny surfaces?	0	1	2	3
10. Maintaining focus for prolonged near work?	0	1	2	3

FINALLY	**Completely Satisfied**	**Very Satisfied**	**Very to Moderately Satisfied**	**Moderately to A Little Satisfied**	**A Little Satisfied**	**Completely Unsatisfied**
Overall how satisfied are you with the near visual ability that you have?	0	1	2	3	4	5

THIS IS THE END OF THE QUESTIONNAIRE. THANK YOU FOR YOUR TIME.

Appendix 6 - American Academy of Optometry poster presented in Nov 2018

Aston University
Birmingham

ORG Ophthalmic Research Group

Reliability and repeatability of the reading performance metrics on different reading charts in pre-presbyopic and presbyopic subjects

Mashaaer Baashen, Shehzad A. Naroo, Frank Eperjesi
Email: baashenm@aston.ac.uk

جامعة الملك سعود
King Saud University

Introduction

Reading difficulty is one of the most common reason for people presenting for an eye examination, as the reading ability has the largest impact factor on a person's daily life and functioning (Fujita et al., 2008). So, to understand how near vision is measured and which tests are done to investigate it is important. In clinical practice and in research, reading tests may randomly chosen to assess reading performance. However, to obtain comparable outcomes, a reliable and repeatable reading test chart must be chosen.

In contrast to distance visual acuity, there is no agreement regarding the charts or tests that should be used for measuring near reading acuity and reading performance, resulting in unstructured development of many different reading charts.

Studies describing reading test chart measurement properties (repeatability and reliability) are lacking. Also, the content and the quality of the study design, scoring rule methods and analytic methods used are so varied, making it difficult to compare them.

Purpose

The purpose of this study was to compare the inter-chart reliability and test-retest repeatability of six different reading charts in the English language (MNread, Radner, Colenbrander, IReST, Baily-Lovie and Times New Roman) for measuring reading performance metrics with respect to reading acuity (RA), critical print size (CPS), maximum reading speed (MRS) and LogRAD/LogMAR ratio in pre-presbyopic and presbyopic subjects by using the same study design and algorithm for scoring the reading performance metrics in each chart.


MNread


Radner


Colenbrander


IReST


Bailey-Lovie


Times New Roman

Methods

At two testing sessions held at two to three weeks apart the data were collected from

29 pre-presbyopic
Age range from 18 to 31 years

29 presbyopic
Age range from 41 to 70 years

- All the participants were native English speakers with no ocular disease or binocular anomalies that may influence the cognitive reading ability
- All reading tests were presented on a reading stand at a distance of 40cm. To ensure a constant viewing distance, a headrest for the forehead was used. Baily-Lovie word chart is designed to be viewed at 25cm. However, in this study, the Bailey-Lovie reading chart was presented at 40 cm and the correction distance factor was used
- The participants read all of the sentences in the reading charts, loudly accurately, as quickly as possible in a randomized order using one eye only.
- The reading performance metrics outcomes were described as follows: reading acuity (RA), maximum reading speed (MRS), critical print size (CPS) and logMAR/logRAD ratio.
- For the IReST reading test, just one metric was measured, which was the average reading speed, because the IReST test is a one print size chart.

Results

Repeatability results of the reading performance metrics between the two visits in pre-presbyopic subjects

Reading chart	Reading performance metrics	Correlation coefficient (r value)	Differences in measurements (t test) p value	Bland-Altman agreement
Colenbrander	RA (logRAD)	0.772	0.572	Low *
	MRS (wpm)	0.550	0.920	High
	CPS (logMAR)	0.297	0.206	High
	LogRAD/LogMAR ratio (%)	0.755	0.239	High
Radner	RA (logRAD)	0.630	0.008 *	Low*
	MRS (wpm)	0.895	0.01 *	Low *
	CPS (logMAR)	0.424	0.687	High
	LogRAD/LogMAR ratio (%)	0.582	0.022	High
MNread	RA (logRAD)	0.816	0.593	High
	MRS (wpm)	0.856	0.291	High
	CPS (logMAR)	0.509	0.148	Low
	LogRAD/LogMAR ratio (%)	0.701	0.257	High
Bailey-Lovie	RA (logRAD)	0.795	0.001*	High
	MRS (wpm)	0.572	0.172	High
	CPS (logMAR)	0.580	0.244	Low
	LogRAD/LogMAR ratio (%)	0.608	0.003	High
Times New Roman	MRS (wpm)	0.687	0.094	Low*
	CPS (N notation)	0.272	0.169	Low*
IReST	Average reading speed (wpm)	0.918	0.0002*	High

* Statistical significant differences or low agreement in the reading performance measurements between the two testing session which indicate poor test-retest repeatability of the chart

In the Times New Roman chart, all the pre-presbyopic participants could easily read the smallest print size, N6, in the two testing sessions. Therefore, the reading acuity was exactly the same in the two visits (N6)

The MNread chart showed a very high repeatability for measuring RA, MRS and LogRAD/LogMAR ratio without any significant differences between the two testing sessions

Repeatability results of the reading performance metrics between the two visits in presbyopic subjects

Reading chart	Reading performance metrics	Correlation coefficient (r value)	Differences in measurements (t test) p value	Bland-Altman agreement
Colenbrander	RA (logRAD)	0.620	0.096	Low*
	MRS (wpm)	0.560	0.673	High
	CPS (logMAR)	0.599	0.477	High
	LogRAD/LogMAR ratio (%)	0.746	0.089	High
Radner	RA (logRAD)	0.805	0.007*	High
	MRS (wpm)	0.781	0.028*	High
	CPS (logMAR)	0.638	0.023*	High
	LogRAD/LogMAR ratio (%)	0.807	0.009*	High
MNread	RA (logRAD)	0.845	0.551	High
	MRS (wpm)	0.767	0.441	High
	CPS (logMAR)	0.551	0.284	High
	LogRAD/LogMAR ratio (%)	0.420	0.252	Low*
Bailey-Lovie	RA (logRAD)	*0.574*	0.489	High
	MRS (wpm)	*0.713*	0.621	High
	CPS (logMAR)	*0.550*	0.757	High
	LogRAD/LogMAR ratio (%)	*0.645*	0.571	High
Times New Roman	MRS (wpm)	*0.818*	0.969	High
	CPS (N notation)	0.430	0.663	Low *
IReST	Average reading speed (wpm)	0.941	0.880	High

The Bailey-Lovie reading chart showed a good repeatability results for measuring all the reading performance metrics without any significant differences between the two testing sessions in presbyopic subjects

Again for times New Roman chart, the reading acuity for all the presbyopic participants was N6. This findings confirmed that Times New Roman chart has the least reliability for measuring RA.

Reliability results of the reading performance metrics in pre-presbyopic subjects

The reading chart show good reliability results for reading performance metrics measurements compared to other reading charts In pre-presbyopic subjects was **Bailey-Lovie reading chart**. the reliability analysis for RA (α = 0.881), MRS (α =0.772), CPS (α = 0.717) and LogRAD/LogMAR (α =0.750) ratio resulted in good Cronbach's alpha coefficients


Bailey-Lovie reading chart

Reliability results of the reading performance metrics in presbyopic subjects

The reading chart show good reliability results for reading performance metrics measurements compared to other reading charts In pre-presbyopic subjects was **MNread** reading chart. the reliability analysis for RA (α = 0.915), MRS (α =0.867), CPS (α = 0.710) and LogRAD/LogMAR (α =0.893) ratio resulted in good Cronbach's alpha coefficients


MNread reading chart

Conclusion

The use of standardised reading test for measuring reading performance metrics is of significant scientific and clinical value. This study confirmed that the clinical results of the reading performance metrics obtained with different reading charts are affected by the choice of the chart.

For pre-presbyopic subjects, the test-retest repeatability for measuring reading performance metrics are higher in **MNread** chart. While **Bailey-Lovie** word chart has a higher inter-chart reliability.

For presbyopic subjects, the test-retest reliability for measuring reading performance metrics are higher in **Bailey-Lovie** word chart. While **MNread** chart has a higher inter-chart reliability.

Taken together, from these results, the authors suggest the use of MNread chart for all patients of any age group should be used if only one reading test chart is required. As the MNread chart does not revealed any statistical significant differences for measuring all reading performance metrics in the pre-presbyopic and presbyopic participants.

This finding has important implications for evaluating the reading performance in clinical practice or scientific research. Also, this study will be interest when evaluating the repeatability and reliability of reading test charts in some groups of patients where the measuring of near vision is more crucial such as patients with cataract or low vision.

References

BAILEY, I. L. & LOVIE, J. E. 1976. New Design Principles for Visual Acuity Letter Charts". *Optometry & Vision Science*, 53, 740-745
FUJITA, K., ODA, K., WATANABE, J. & YUZAWA, M. 2008. How normal eyes perform in reading low-contrast texts. *Japanese journal of ophthalmology*, 52, 44-47