Repeatability and Reproducibility of Buari-Chen Malay Reading Chart

Nursyairah Mohd Khalid and Ai-Hong Chen*

Optometry, Faculty of Health Sciences, Universiti Teknologi MARA,
Cawangan Selangor, Kampus Puncak Alam, Malaysia

**Corresponding author's email: aihong0707 [AT] yahoo.com*

ABSTRACT— ***The purpose of our study was to evaluate the repeatability and reproducibility of Buari-Chen Malay Reading Chart (BCMRC). Thirty normally sighted young adults (mean age: 20.6 years) were recruited based on convenient sampling. They were instructed to be read aloud 2 sets of BCMRC (contextual sentences (CS) set and random words (RW) set) at random sequence in three different sessions. The repeatability limits of the reading speed for contextual sentences (CS) set and random words (RW) set were 45.46 wpm and 29.16 wpm respectively, while the limit for reproducibility was 33.12 wpm for CS set and 28.38 wpm for RW set. Bland and Altman's plot showed good agreement for both sets of BCMRC as most differences of reading speed between reading sessions were placed within the limits of agreement (LoA). The mean difference in repeatability for CS set was 13.14 wpm (95% LoA = 40.81 to -14.53) and 8.96 wpm (95% LoA = 32.63 to -14.70 for RW set. For reproducibility, the mean difference was 11.30 wpm (95% LoA = 36.38 to -13.78) for CS set and 2.79 wpm (95% LoA = 31.13 to -25.55) for RW set. The BCMRC showed good repeatability and reproducibility that supported its usage as reading research tool as well as clinical diagnostic tool.***

Keywords— BCMRC, Precision, Repeatability, Reproducibility

1. INTRODUCTION

Measuring the reading performance is one of the important components in assessing visual function. The performance-based reading test provides better discrimination in reading ability level than self-report [1]. The influx of reading tools has been widely studied and improved for clinical and experimental purpose [2]. There are nine properties listed as reading performance assessment tools for three standard continuous reading tests (MNread, Radner, and IReST), which are internal consistency, content validity, inter-chart reliability, test repeatability, test reproducibility, inter-rater and intra-rater reliability, cross-cultural validity, and generalizability [3].

The combined features of contextual sentences and random words reading chart in the Malay language known as Buari-Chen Malay Reading Chart (BCMRC) [4], [5] was designed to be used among Malay native speakers to eliminate the language barrier while assessing the reading performance among the population. Even though some developed standardize reading charts were translated into other languages such as Italian [6], Spanish [7], Turkish [8], Dutch [9], and Japan [10], each developed chart had their own unique designs and reasons of development. The BCMRC was reported to have good internal consistency and high reliability when compared with standardized English reading chart; MNread and Bailey-Lovie [4], [5]. Several studies regarding the measurement properties of the reading performance assessment tools were conducted on a developed standardized reading charts such as MNread acuity chart [6], [11], [12], Radner reading chart (RRC) [9], [13], [14], and IReST [15]. As mentioned by Brussee (2014), MNread acuity chart showed good research on repeatability while RRC had good reliability and reproducibility. However, repeatability and reproducibility are yet to be reported for BCMRC.

Repeatability and reproducibility are two components of precision of clinical device or clinical method [16]. Precision refers to the closeness between repeated measurements [17]. In measuring the precision for reading materials, repeatability is the variability obtained in repeated measurements by controlling other factors (i.e. methods, conditions) [3]. The contributing factors comprised of examiners, reading charts used, and time interval between measurements. Reproducibility referred to the variability obtained in repeated measurement with one or more factors that were varied such as conditions (i.e. time, rater) or methods (i.e. chart, test) [3]. There were various statistical methods used in reporting precision (repeatability and reproducibility) measurement for reading tests, as shown in Table 1. Reporting of the correlation coefficient for precision was questionable as it described the closeness of the linear relationship. Hence, it

should be avoided due to certain limitations as well as difficulties to detect a systematic error [3], [16]. Thus, the approach of limits of agreement proposed by Bland and Altman was the preferred method and should be applied as appropriate statistical methods for the precision measurement [3].

Table 1: Summary of the repeatability and reproducibility of published reading charts

Charts	Sources	PCC	ICC	S_r, r	S_R, R	B&A
MNread	[6]	0	+	0	0	+
RRC	[14]	+	0	0	0	+
MNread	[12]	0	0	0	0	+
RRC	[9]	0	0	+	+	0
RRC	[18]	0	+	+	+	0
MNread	[19]	0	+	0	+	0
UiTM-Mrw	[20]	0	0	0	0	+
RRC	[21]	+	0	0	0	0
RRC	[22]	+	0	0	0	0
UiTM-Muw	[23]	0	0	0	0	+

+, reported; 0, not reported; PCC, Pearson correlation coefficient; ICC, Intra-class correlation coefficient; S_r, repeatability; r, repeatability limit; S_R, reproducibility; R, reproducibility limit; B&A, Bland and Altman; MNread, MNread acuity chart; RRC, Radner Reading Chart; UiTM-Mrw, UiTM Malay related words; UiTM-Muw, UiTM Malay unrelated words.

Most studies on reading test reported the findings in term of the value of correlation and advance with Bland and Altman's plot. Test-retest analysis for RRC among macular diseases patients that was conducted for three to four weeks of interval between sessions reported high correlation for all three charts [14]. There was no significant difference in maximum reading speed (MRS) between two repeated sessions ($p = 0.12$) and Bland and Altman's plot also evaluated a good reproducibility [14]. High test-retest reliability for the MNread acuity charts was found among children [6]. The study reported high intraclass correlation coefficient (ICC) value of maximum reading speed (MRS), ICC = 0.95, with 95% confidence limits of agreement ±0.077 logWPM, which corresponds to ±19% change in wpm [6]. Repeatability of MNread among young adults population [12] showed no practice effects, with a mean difference (95% CI) of 1.4 wpm (9.96 to -7.16) and coefficient of repeatability of ±8.6 wpm for reading speed. Validation of MNread in Greek version illustrated good ICC for MRS = 0.87, with a coefficient of repeatability 46.96 wpm [19]. All three versions proved to be reproducible with no negligible between-sessions and between-charts variability [19]. Every standardized reading chart was suggested to fulfill the reading chart measurement properties in order to qualify and be applicable in the clinical and research settings. Our study was conducted to establish the repeatability and reproducibility to further investigate the precision of BCMRC in order to qualify its usage in clinical and research settings. Our findings filled up gaps on the precision aspect of BCMRC. The agreement between sessions was aimed to determine if BCMRC might be used repeatedly and interchangeably.

2. MATERIALS AND METHODS

Experimental cross-sectional study design was executed to evaluate the repeatability and reproducibility of BCMRC. BCMRC consisted of four charts in Malay language with division into two sentence arrangement; contextual sentences (CS) set and random words (RW) set [4], [5]. Each of the charts consists of 14 sentences with six words in each sentence, in different size range; from 8.0M to 0.4M equivalent to 1.3 logMAR to 0.0 logMAR. The chart was printed in Arial font type. Each print size consists of 12 syllables, 31 to 35 characters including spaces. The reading chart was placed at 45° on the reading stand and the reading distance was set at 40 cm. Each sentence was covered with a white blank card to avoid pre-reading. All subjects were assigned to read both CS and RW sets in random order, loudly and accurately as fast as possible from the largest to the smallest line. The complete reading procedures were recorded using the voice recorder and were heard during data extraction. The time taken to read each sentence was recorded and any error made during reading was noted and classified [24], [25]. Measurement of precision in our study was measured via repeatability and reproducibility.

2.1 Subjects Selection

Thirty young adults (mean age: 20.60±0.81 years) with normal vision and fluent in oral and reading of Malay language reading materials were recruited. Informed consent was obtained. Our study adhered to the tenets of Declaration of Helsinki and was approved by the Research Ethic Committee of the University (Approval Code: 600-RMI (5/1/6) REC/108/15). Subjects had no previous history of ocular diseases or received any medication that could influence the outcomes of the reading performance. All subjects wore their habitual distance correction and had distance visual acuity of at least 0.1 logMAR or better binocularly using Bailey-Lovie distance visual acuity charts.

2.2 Repeatability and Reproducibility

The repeatability was tested by repeating the measurements in two separate occasions with 24-hour break in between using the same examiner and scheduling for the same visiting hour as the initial session. Time between each session should be the shortest possible period for repeatability [17]. On the other hand, reproducibility was tested by repeating the reading procedures with two different examiners. Good agreements of reading speed measurement between two examiners provided an estimation of reproducibility. A minimum of 2-hour gap was given to each subject in between sessions. The two examiners followed the same instruction sheet as developed in our study.

2.3 Data Analysis

Data was analyzed using Statistical Package for Social Sciences (SPSS) version 20.0. The main parameter calculated was reading speed, in words per minute (wpm). The reading speed was determined for each sentence by calculating the number of correct words divided by the time taken to read the sentence [12]. An independent-samples *t*-test was conducted to test for equality of means of reading speed obtained in the initial and repeated session for repeatability and reproducibility. Within-subject standard deviation was measured using a simple one-way analysis of variance (ANOVA). The square root of the mean square within groups was calculated to provide the value of repeatability (S_r) and reproducibility (S_R). The value of the limit for repeatability (r) and reproducibility (R) was considered at 95% of measurements of S_r and S_R that should occur, given the equations $1.96\sqrt{2}$ x S [17].

Bland and Altman's plot were used to assess the agreement between two variables for both repeatability and reproducibility. The agreement was summarized by determining the bias, which was estimated by the mean difference, *d*, and the standard deviation of the differences, *s* [16]. The graph information was plotted with differences against the means [16]. The limits of agreement (LoA) described the difference of reading speed between sessions ($d \pm 1.96s$). The 95% confidence intervals of upper and lower limits of agreement were also calculated and plotted in the graph using the equation of LoA ± (*t* distribution x standard error). The *t* value was 2.05, where the 95% confidence intervals for this study was 29 degrees of freedom (n - 1) and the standard error was calculated from the formula $\sqrt{3s^2/n}$ where n is the sample size.

3. RESULTS

3.1 Repeatability

The Buari-Chen Malay Reading Chart (BCMRC) showed good repeatability of reading speed (wpm). The mean and standard deviation for two separate sessions of contextual sentences (CS) set and random words (RW) set of Buari-Chen Malay Reading Chart (BCMRC) were summarized in Table 2. The initial and repeated means of reading speed were not significantly different for both sets of chart, $p > 0.05$.

Table 2: Summary of BCMRC repeatability findings for both CS set and RW set

	Contextual sentences set	Random words set
Mean initial reading speed (SD)	194.42±26.41 wpm	145.96±22.15 wpm
Mean repeated reading speed (SD)	207.56±26.37 wpm	154.93±17.54 wpm
Mean difference (LoA)	13.14 (40.81 to -14.53)	8.96 (32.63 to -14.70)
Independent samples *t*-test	$t(58) = -1.93$, $p = 0.06$	$t(58) = -1.74$, $p = 0.08$

Notes: Mean values with standard deviation (SD) and mean difference with a limit of agreement (LoA) in parentheses of reading speed in word per minute (wpm) for contextual sentences set and random words set on two separate testing sessions of repeatability.

The repeatability (S_r) of the reading speed for CS set was 16.40 wpm with repeatability limit (r) of 45.46 wpm. For RW set, the repeatability, S_r, was 10.52 wpm with a limit, r, of 29.16 wpm. Figure 1 and Figure 2 showed that difference of up to 13.14 wpm between repeated sessions for contextual sentences set and 8.96 wpm for random words set respectively. The 95% limits of agreement between repeated sessions were ±1.44 logWPM for contextual sentences set and ±1.37 logWPM for random words set. Most differences were within the limits of agreement, considering differences followed a normal distribution. The difference of reading speed for both contextual sentences and random words sets was small enough to support clinical application.

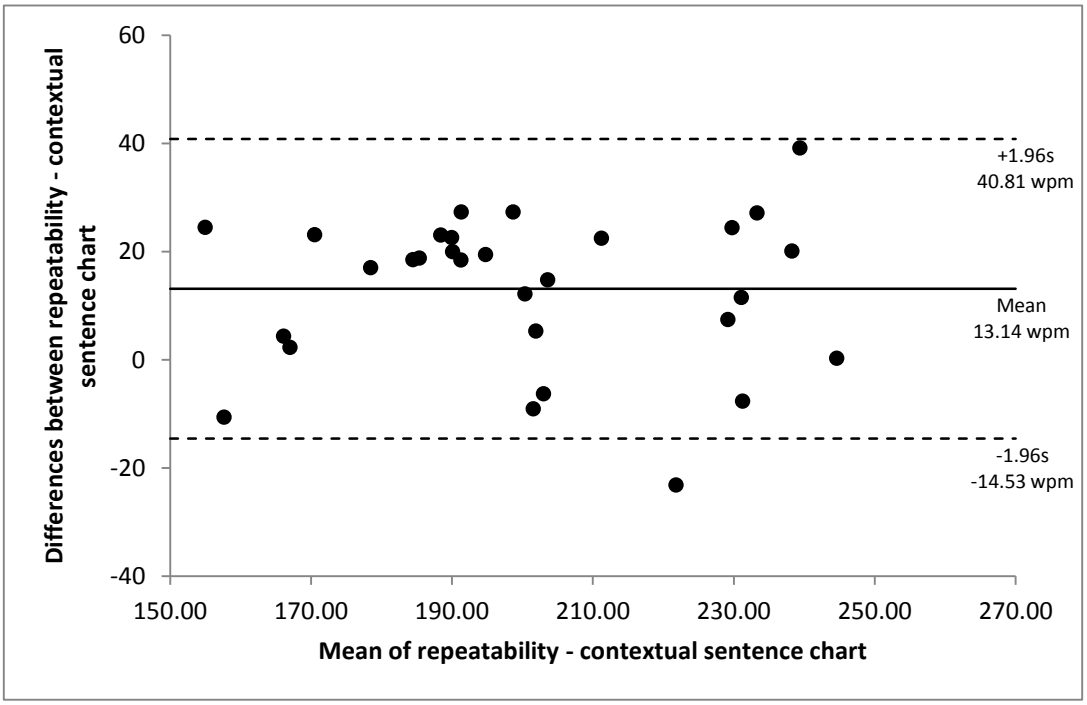

Figure 1: Bland and Altman plot for repeatability evaluation of the contextual sentences set. The solid black line showed the mean difference, the limits of agreement by the upper and lower dotted lines and the 95% CIs.

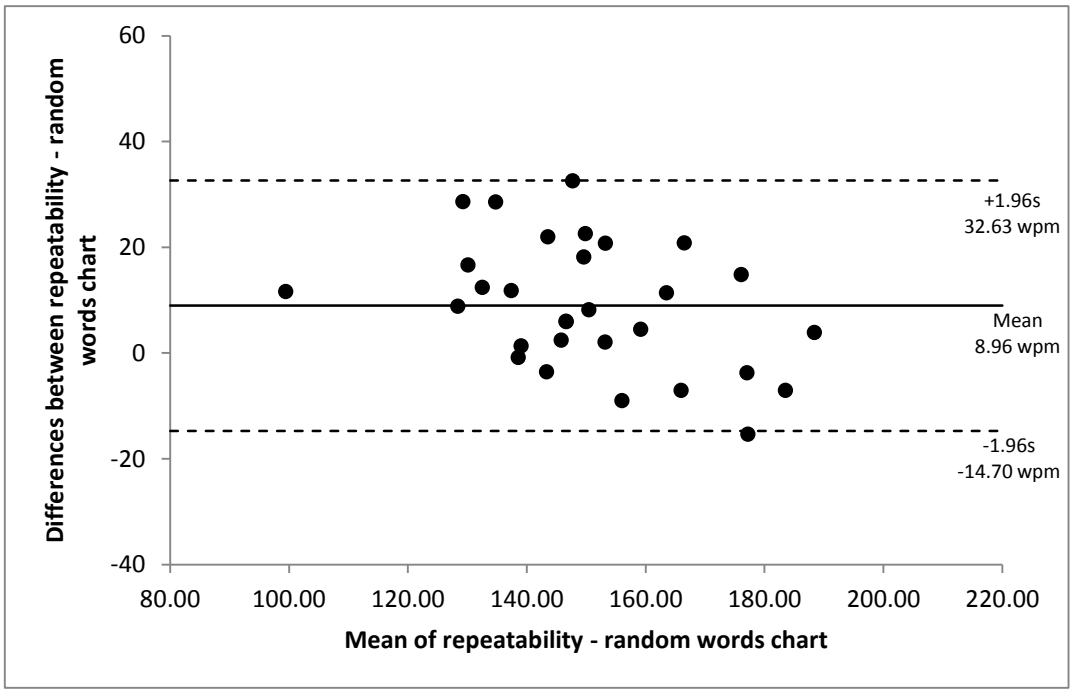

Figure 2: Bland and Altman plot for repeatability evaluation of the random words set. The solid black line showed the mean difference, the limits of agreement by the upper and lower dotted lines and the 95% CIs.

3.2 Reproducibility

Both sets of BCMRC illustrated a good inter-examiner reproducibility of reading speed. The mean and standard deviation for reading speed obtained on two separate testing sessions assessed by two different examiners were tabulated in Table 3. An independent-samples *t*-test for reproducibility also showed no statistically significant difference in reading speed attained by the first and second examiner. The calculated value of reproducibility (S_R) for contextual sentences (CS) set and random words (RW) set were 11.95 wpm and 10.24 wpm respectively and the values of the limit of reproducibility (R) were 33.12 wpm for CS set and 28.38 wpm for RW set.

Table 3: Summary of BCMRC reproducibility findings for both CS set and RW set

	Contextual sentences set	Random words set
Mean reading speed by the 1st examiner (SD)	194.42±26.41 wpm	145.96±22.15 wpm
Mean reading speed by the 2nd examiner (SD)	205.71±22.44 wpm	148.75±20.31 wpm
Mean difference (LoA)	11.30 (36.38 to -13.78)	2.79 (31.13 to -25.55)
Independent samples *t*-test	$t(58) = -1.79$, $p = 0.09$	$t(58) = -0.51$, $p = 0.61$

Notes: Mean values with standard deviation (SD) and mean difference with a limit of agreement (LoA) in parentheses of reading speed in word per minute (wpm) for contextual sentences set and random words set on two separate testing sessions of reproducibility.

Using Bland and Altman analysis (Figure 3 and Figure 4), the mean difference of reading speed for CS set and RW set tested by different examiners were 11.30 wpm and 2.79 wpm, respectively with LoA 36.38 wpm to -13.78 for CS set and 31.13 wpm to -25.55 wpm for RW set. The 95% confidence limit of agreement for CS set was ±1.40 logWPM and for RW set is ±1.45 logWPM. Most differences between sessions of reproducibility were laid within the LoA for both sets of BCMRC. Reading speed was not influenced by the differences obtained between different examiners.

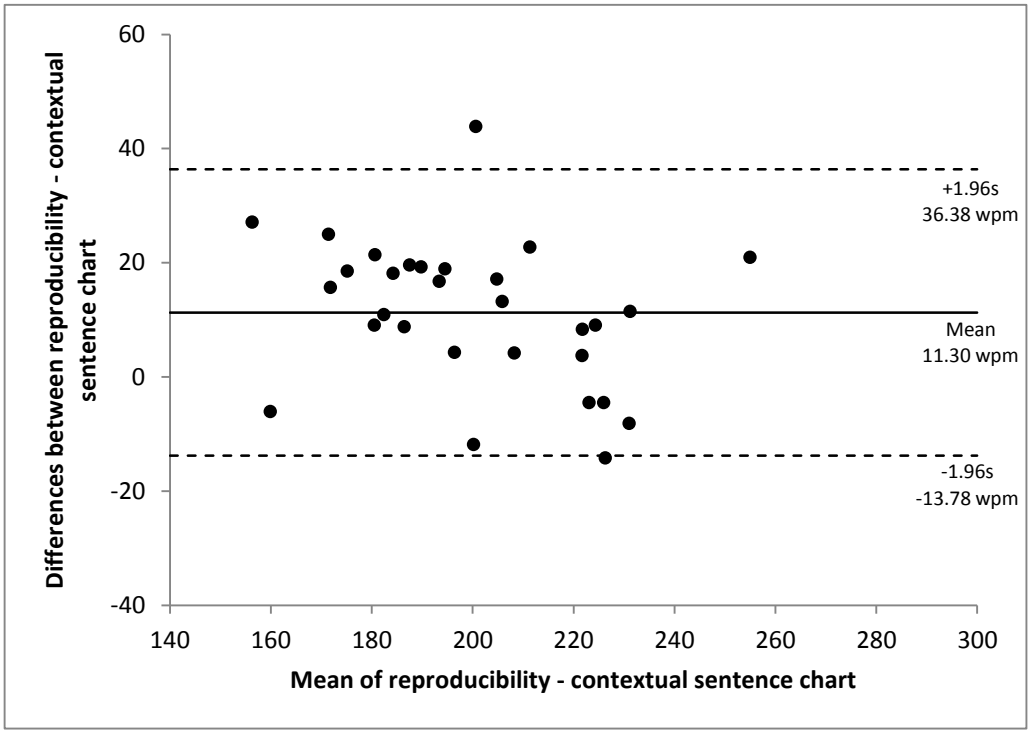

Figure 3: Bland and Altman plot for reproducibility evaluation of the contextual sentences set. The solid black line shows the mean difference, the limits of agreement by the upper and lower dotted lines and the 95% CIs.

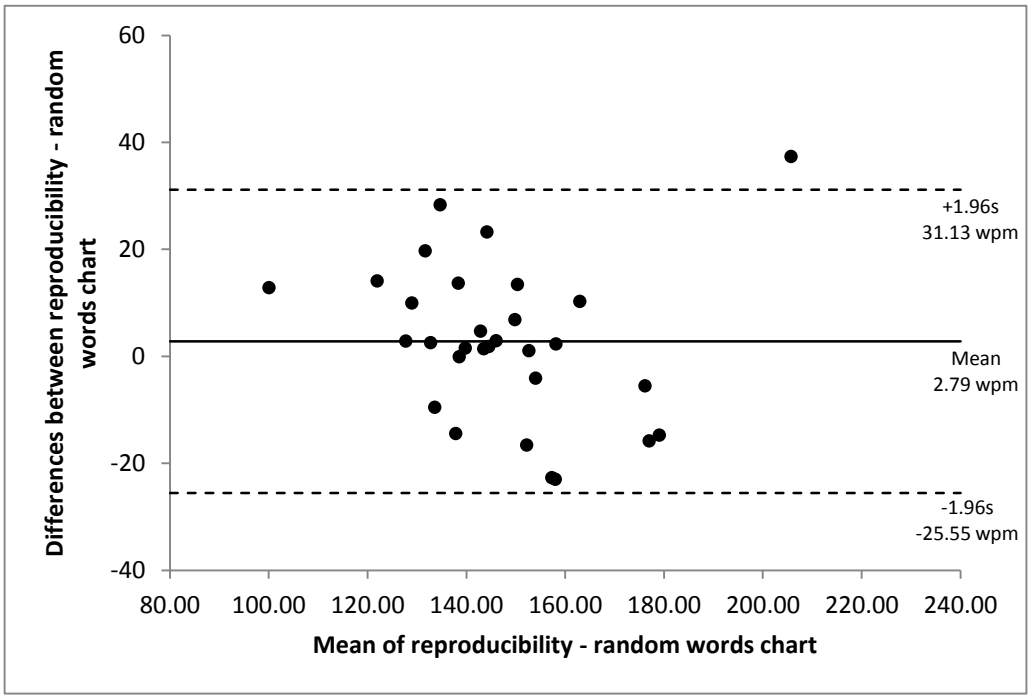

Figure 4: Bland and Altman plot for reproducibility evaluation of the random words set. The solid black line shows the mean difference, the limits of agreement by the upper and lower dotted lines and the 95% CIs.

4. DISCUSSION

The two sets of BCMRC, contextual sentences (CS) set and random words (RW) set, provided good precision as a result of testing on repeatability and reproducibility. Both assessments showed no statistically significant difference when tested using independent-samples *t*-test ($p > 0.05$). The findings of our study were consistent with most previous reports on repeatability and reproducibility of well-developed reading tools. Repeatability and reproducibility of MNread acuity charts showed no significant difference in reading speed measurement that calculated in three different testing distances [12]. The validation of the Greek MNread acuity chart tested on normally sighted adults gave the repeatability limit of 46.96 wpm [19], while our study showed the repeatability limit for CS set was 45.46 wpm. Furthermore, the feasibility of Radner Reading Chart (RRC) in low vision patients showed similar value of repeatability, 42.53 wpm [18] compared with our study even though the age and the visual status were disparate. However, the Dutch RRC showed smaller repeatability limit (22.7 wpm) [9] compared to our study on CS set (45.46 wpm) due to the differences in age and visual status of subjects tested. Older subjects with macular diseases were used in their study while the subjects for our study were young adults with normal vision. Our findings showed that the repeatability of reading speed was twice higher compared to the macular diseases patients [9]. The comparison values were matched with CS set of BCMRC finding as the MNread acuity chart and RRC were both contextual sentences design reading charts. The reproducibility outcomes of reading speed were differed by different factors of variability. Our reading speed of CS set was varied by 33.12 wpm when re-measured with a different examiner. The reading speed of RRC differed by 43.23 wpm after 1-month interval tested among low vision patients [18]. Similar to one month interval, the reading speed among macular diseases patients showed a difference of 28.5 wpm [9]. A moderate repeatability and reproducibility were assumed if there was a considerable variation in repeated measurements from the same subjects [18]. The acceptable level of test-retest repeatability depends on how the test was used and the parameters that were being tested.

Repeatability and reproducibility were also explored using Bland and Altman plot of agreement [12], [14], [20], [23]. Confidence intervals were calculated and reported as the limits of agreement were only the estimation for the whole population [16], [26]. The mean difference of reading speed and their limits of agreement (LoA) for repeatability in our study for both contextual sentences set (16.53 wpm, LoA = 48.96 to -15.90) and random words set (8.96 wpm, LoA = 32.63 to -14.70), was much higher compared to those found by Subramaniam and Pardhan (2006) for repeatability on MNread acuity charts (1.4 wpm, LoA = 9.96 to -7.16). The discrepancies might be due to the time interval between repeated sessions and the characteristics of the chart used. Theirs took a shorter break (exact duration was not mentioned) between repeated sessions [12] while our study chose 24-hour gap before the next session to reduce the learning effect. MNread acuity chart that was used contained 19 sentences, each enclosed with 60 characters [12]. BCMRC used in our study consisted of 14 sentences with 31 to 35 characters each. BCMRC used shorter sentence compared to MNread

acuity chart. The learning effect would be more prominent in reading BCMRC. Direct comparison for mean difference and limits of agreement of reproducibility in our study and previous studies might not be possible due to diverse variables tested such as testing distance [12] and different reading tests used [20], [21], [23].

Mean difference of reading speed for contextual sentences set for both repeatability and reproducibility were found to be higher compared to random words set in our study. The discrepancy might be due to the existence of comprehension by reading the contextual sentences [25], as the sentences were easily remembered. Therefore, whether contextual sentences set or random words set, each has its own advantages and distinct purpose of development. Good agreement of repeatability and reproducibility found in all sets of BCMRC supported its interchangeable use in both clinical and research settings. Future study might explore reading speeds on different age groups and different visual status to investigate additional measurement property of reading chart: generalizability, where the outcomes can be generalized.

5. CONCLUSION

The Buari-Chen Malay Reading Chart provided good repeatability and reproducibility. BCMRC provided a stable range of difference when used repeatedly either under the same condition; highly reliable of test-retest repeatability, or different conditions; highly reliable of inter-examiner reproducibility.

6. ACKNOWLEDGEMENT

This project was supported by Fundamental Research Grant Scheme (600-RMI/FRGS 5/3) (119/2014) from the Ministry of Higher Education Malaysia. We are grateful to Noor Halilah Buari for insightful discussion.

7. REFERENCES

[1] Rubin, G. S. (2013). Measuring reading performance. *Vision research*, *90*, 43-51.
[2] 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.
[3] Brussee, T., Nispen, R., & Rens, G. H. (2014). Measurement properties of continuous text reading performance tests. *Ophthalmic and Physiological Optics*, *34*(6), 636-657.
[4] Buari N. H. and Chen A. H., Buari-Chen Malay Reading Chart (BCMRC): Contextual Sentence and Random Words 2-in-1 Design in Malay. Invention, Innovation & Design Exposition 2017, 25-29 September, Universiti Teknologi MARA Shah Alam, Malaysia. Available at: http://ribu.uitm.edu.my/v2/index.php/27-event/58-iidex2017, iidex2017/B/C05/005.
[5] Buari N. H. and Chen A. H., "Buari-Chen Malay Reading Chart (BCMRC): Contextual Sentence and Random Words 2-in-1 Design in Malay," *Pertanika J. Sci. Techmology*, Special Issue on Interdisciplinary Research, JST. Vol. 25(S). Aug 2017.
[6] Virgili, G., Cordaro, C., Bigoni, A., Crovato, S., Cecchini, P., & Menchini, U. (2004). Reading acuity in children: evaluation and reliability using MNREAD charts. *Investigative ophthalmology & visual science*, *45*(9), 3349-3354.
[7] 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*(4), 638-642.
[8] İDİL, 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*(4), 565-570.
[9] Maaijwee, K., Mulder, P., Radner, W., & Van Meurs, J. C. (2008). Reliability testing of the Dutch version of the Radner Reading Charts. *Optometry and Vision Science*, *85*(5), 353-358.
[10] Ishii, M., Seki, M., Harigai, R., Abe, H., & Fukuchi, T. (2013). Reading performance in patients with glaucoma evaluated using the MNREAD charts. *Japanese journal of ophthalmology*, *57*(5), 471-474.
[11] Mansfield, J. S., Ahn, S. J., Legge, G. E., & 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.
[12] Subramanian, A., & Pardhan, S. (2006). The repeatability of MNREAD acuity charts and variability at different test distances. *Optometry and vision science*, *83*(8), 572-576.
[13] Radner, W., Willinger, U., Obermayer, W., Mudrich, C., Velikay-Parel, M., & Eisenwort, B. (1998). Eine neue Lesetafel* zur gleichzeitigen Bestimmung von Lesevisus und Lesegeschwindigkeit. *Klinische Monatsblätter für Augenheilkunde*, *213*(09), 174-181.
[14] Stifter, E., König, F., Lang, T., Bauer, P., Richter-Müksch, S., Velikay-Parel, M., & Radner, W. (2004). 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*(1), 31-39.
[15] Trauzettel-Klosinski, S., & Dietz, K. (2012). Standardized assessment of reading performance: the new International Reading Speed Texts IReST. *Investigative ophthalmology & visual science*, *53*(9), 5452-5461.

[16] 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*(4), 330-338.
[17] McAlinden, C., Khadka, J., & Pesudovs, K. (2015). Precision (repeatability and reproducibility) studies and sample-size calculation. *Journal of Cataract & Refractive Surgery*, *41*(12), 2598-2604.
[18] Burggraaff, M. C., van Nispen, R. M., Hoek, S., Knol, D. L., & Van Rens, G. H. (2010). Feasibility of the Radner Reading Charts in low-vision patients. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *248*(11), 1631-1637.
[19] 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*(1), 25-31.
[20] Buari, N. H., Chen, A. H., & Musa, N. (2014). Comparison of reading speed with 3 different log-scaled reading charts. *Journal of optometry*, *7*(4), 210-216.
[21] Calossi, A., Boccardo, L., Fossetti, A., & Radner, W. (2014). Design of short Italian sentences to assess near vision performance. *Journal of optometry*, *7*(4), 203-209.
[22] 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*(8), 1297-1303.
[23] Buari, N. H., Azizan, M. F., & Chen, A. H. (2015). Comparison of Reading Speed Using Malay Unrelated Word Reading Chart with Standardized English Reading Charts. *International Journal of Medical and Health Sciences Research*, *2*(3), 55-61.
[24] Douglas, G., Grimley, M., Hill, E., Long, R., & Tobin, M. (2002). The use of the NARA for assessing the reading ability of children with low vision. *British Journal of Visual Impairment*, *20*(2), 68-75.
[25] Khalid, N. M., Buari, N. H., & Chen, A. H. (2017). Comparison of Oral Reading Errors between Contextual Sentences and Random Words among Schoolchildren. *International Education Studies*, *10*(1), 47-55.
[26] Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, *327*(8476), 307-310.