Development and Validation of a New Design Self-assessment logMAR Visual Acuity Test (“Chudjane” iPhone- and iPad-based Application) in a Normal Eyes Population

Nuttawut Rodanant, M.D., Comkrit Mahasith, M.D., Somanus Thoongsuwan, M.D., Nopasak Phasukkijwatana, M.D., Supalert Prakhunhungsit, M.D.
Department of Ophthalmology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

ABSTRACT
Objective: To validate and further design the “Chudjane” application (app), a new design self-assessment logMAR test for distance visual acuity (VA), by comparing the results against the use of a standard numeric ETDRS chart in normal eye population.

Materials and Methods: In total, 52 volunteers who had a normal eye exam and best-corrected VA score by numeric ETDRS (NE) chart equal to or better than 6/6 (logMAR score 0.00 or less) were included. The “Chudjane” app with 3 patterns of optotypes (Arabic numbers (AN), Tumbling-E (TE) and Landolt-C (LC)) was used twice to assess VA individually.

Results: The mean VA in each test NE, AN, TE, LC from the first round were -0.06, -0.10, -0.08 and -0.04, respectively compared to -0.07, -0.12, -0.09 and -0.05 from the second round respectively. Comparing results from the first and second round revealed that NE and LC had higher test-retest reliability (ICC=0.712, 0.789 respectively) than AN and TE (ICC=0.140, 0.495 respectively). For validity, result from NE was compared to each app test using the second round values. Modified Bland-Altman plot showed the mean differences (95% LOA) for NE-AN, NE-TE and NE-LC of 0.05 (-0.11 to 0.20), 0.02 (-0.11 to 0.15) and -0.03 (-0.19 to 0.13) respectively. Simple linear regression analysis of the difference (i.e. NE-AN, NE-TE and NE-LC) on NE showed that the difference did not depend on the NE value with slope close to zero.

Conclusion: The study demonstrated that by using the «Chudjane» application, LC had higher test-retest reliability and higher validity than TE and AN compared to the standard ETDRS chart.

Keywords: Self-assessment; visual acuity; mobile application (Siriraj Med J 2022; 74: 590-599)

INTRODUCTION
Visual acuity is the most commonly performed measurement of visual function in clinical practice. It is used to establish the need for a full evaluation of visual function and to inform the clinical decisions of ophthalmologists.1

Several optotypes are used in clinical practice, such as Arabic numbers (AN), Landolt-C (LC), and Tumbling-E (TE). Landolt-C and Tumbling-E are commonly used for assessing the vision of children, as well as for illiterate and non-English-speaking people; the results are determined by four orientations of the letters.2-4 Landolt-C has been...
adopted as the primary standard optotype for visual acuity testing. Arabic numbers (AN) are usually used in the clinical setting because patients are more familiar with this optotype, which comprise the numbers 0-9, which are typically recognized unequally.

Snellen’s chart is the most common method for assessing visual acuity, mostly because it is easy and quick to use. However, it has some limitations, such as the nongeometric progression of the optotype size from line to line and the variable numbers of letters on each line.

The ETDRS chart is the best-known logMAR chart, and is also used for most major research studies. It overcomes the limitations of Snellen’s chart, as it incorporates geometric progression (step down of 0.10 log units from line to line) and equal numbers of letters on each line (5 letters per line). For the tests, each character is valued at 0.02 logMAR units per letter that is read correctly.

The “Chudjane” test application (app) was designed such that it could be utilized by a user to self-assess their personal visual acuity and the app automatically summarized and reported the results of the visual acuity in logMAR units, together with the duration of the test.

We measured Visual Acuity at “distance” 4 meters (same as standard ETDRS) using both iPad & iPhone together, iPad as a chart at distance and iPhone as a remote control connected to iPad via Bluetooth that can select optotypes, characters and sizes as described in methods.

The limitations of standard ETDRS chart were memorization effect, patients could remember the fix characters on light box chart, and ETDRS character size changed in steps of 0.1 logMAR. This app was designed to prevent the memorization effect, and to be more fine-scale than the standard ETDRS chart.

In Thailand, like in most countries, smartphone use has been increasing every year (69.6% of the total population possessed a smartphone in 2018). The development of healthcare applications has been increasing and there are now more than 100 visual acuity test applications available; albeit only a small number of them have been validated.

Perera et al. performed a study in 2015 on the reliability and accuracy of 11 visual acuity applications on smartphones and reported that their accuracy varied from about 4.4%–39.9%.

In the present study, the “Chudjane” test app was compared with a standard numeric ETDRS chart for assessing its validity agreement, reliability and duration of the test.

**MATERIALS AND METHODS**

**Study participants**

In total, 52 participants who were over 18 years old, healthy, Thai, could read Arabic numbers and who had no ocular disease by a slit lamp biomicroscope examination (except a refractive error that could be corrected by glasses to get best corrected visual acuity (BCVA) equal to or better than 6/6 (logMAR score 0.00 or less) which is the standard for normal eye visual acuity) were included in the study. The exclusion criteria included those who could not cooperate due to physical or mental disease and those who had a history of cycloplegic drugs or anesthetics allergy.

We validated the sample size based on an equivalence study (z-test), which indicated we required a sample size of 45 eyes to achieve 95% power with a significance level of 0.05 in order to detect an equivalence limit difference of 0.10 logMAR. We chose to use 0.10 logMAR because it represents a significant difference (95% confidence interval) for test–retest variations. The standard deviation of difference was 0.20 logMAR, which was based on Bastrawrous et al.’s study. Lastly, an allowance of 15% was made for drop-outs, so the calculated sample size was 52 eyes.

\[
n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\delta - |\mu - \mu_0|)^2}
\]

\(n = \) Sample size  
\(\alpha = \) Type-I error = 5% (Z = 1.645)  
\(\beta = \) Type-II error = 5% (Z = 1.645)  
\(\sigma = \) Standard deviation of differences  
\(\delta = \) Effect size \((|\mu - \mu_0|/\sigma)\)  
\(\mu - \mu_0 = \) Equivalence limit of differences

**Ethics**

The cross-sectional study adhered to the tenets of the declaration of Helsinki and was approved by the Ethics Committee of the Siriraj Institutional Review Board (COA no. Si 696/2018). Informed consent was obtained from all participants. The objectives of the study, examination process, benefits, and risks of the study were explained to all participants. All the participants gave their signed consent to participate.

**“Chudjanes” test app**

The app was written for the iOS platform, and this study used an iPad Pro 10.5” and iPhone 7 Plus for the testing. (Fig 1) The iPad was the main display device
and displayed the optotypes located at 4 meters from the participant being tested. The iPhone served as a remote controller that was connected to the iPad via Bluetooth. (Fig 2) The brightness of the display on both devices was set to 100%.

The “Chudjane” test app was developed based on the logMAR chart, and was designed such that it could be utilized by a user to self-assess their personal visual acuity.

Three types of optotypes (Arabic numbers, Tumbling-E, and Landolt-C) were used and these could be selected as different options.

Arabic numbers (AN) consisted of the numbers 0 to 9, which were ordered in the same way as with a standard numeric ETDRS chart.

The Tumbling-E and Landolt-C optotypes consisted of 4 orientations of E or C (0°, 90°, 180°, and 270°).

This app was designed so that it could randomize all the characters of each optotype to prevent the memorization effect.

The size of the optotypes on the iPad could be controlled by the iPhone and the character size changed in steps of 0.1 logMAR for the Arabic numbers (AN) and 0.01 logMAR for the Tumbling-E (TE) and Landolt-C (LC) characters (allowing TE and LC to be more fine-scale than the standard ETDRS chart (NE) and AN by a factor of up to 10 times).

The app started with a single character displayed on the iPad, which equaled 1.0 logMAR on the ETDRS chart. The test subject used the iPhone to control and
adjust the character size on the iPad display until it was the smallest size that they could still read, and then recorded the result on the app.

For the AN optotype, the app automatically selected the character size that matched the smallest character size in the test subject’s records. Then it selected 5 numbers the same as on the Standard ETDRS chart to make 1 test set (equal to 1 line of the ETDRS).

For Tumbling-E (TE) and Landolt-C (LC) optotypes, the app used the smallest size that was recorded in the test subject’s records and randomly selected 5 characters for each test set. The test subject selected the answer on the iPhone that they believed matched the character on the iPad. After the first test set was completed, the results were automatically analyzed and the system response depended on the following 2 conditions.

In the first condition, if the test subject chose the correct answer for more than 2 out of the 5 characters (the subject got ≥3 correct) in each test set, the size of the 5 characters in the next test set was automatically reduced by 1 step. This condition was repeated until the correct answers in the last test set were less than three out of the 5 characters (the subject got ≤2 correct) or reached -0.3 logMAR. Then, the test was finished.

In the second condition, if the test subject chose the correct answer for less than 3 out of the 5 characters (the subject got ≤2 correct) in each test set, the size of the 5 characters in the next test set was automatically enlarged by 1 step. This condition was repeated until the correct answers in the test set were more than two out of the 5 characters (the subject got ≥3 correct). Then the character size in the last test set was reduced by 1 step, and the test was finished.

After the test subject had finished all the tests, the app automatically summarized and reported the results of the visual acuity in logMAR units, together with the duration of the test.

Scoring

The app scored the test subjects in 2 parts.

Part 1 Arabic numbers (AN): The score was 0.1 logMAR for each test set (5 characters), which equaled 1 line of the standard ETDRS chart. Each character was valued at 0.02 logMAR.

Part 2 Tumbling-E (TE) and Landolt-C (LC): Each character was valued at 0.002 logMAR because the progression of the optotypes was 0.01 logMAR for each test set (5 characters).

Testing protocol

1. All the participants underwent distant visual acuity testing using a back-illuminated 4-meters numeric ETDRS (NE) chart with the same chart and in the same environment. For all the tests, the presenting acuity was measured with the usual eyesight correction if worn. The duration of the test was also recorded.

2. The test subject’s eyes were completely examined by a slit lamp biomicroscope to detect any abnormalities.

3. All the test subjects were given a demonstration on how to use the app.

4. The test subjects used the app to test their visual acuity by themselves.

5. After they had finished the first test (first round of testing) they were allowed to rest for about 15 minutes, and then they were tested again (second round of testing).

6. The test subjects’ fundus was examined.
Statistical analysis

All the visual acuity measurements were converted to logMAR units, and the duration of the tests was converted to seconds. Data were summarized as mean ± SD. Test-retest reliability of NE, 3 App results (i.e., AN, TE and LC) from the first and second round was assessed via scatter plot, intraclass correlation coefficient (ICC, using a 2-way random effect model, absolute agreement and single measure) and classic Bland-Altman plot (X-axis = average from the first and second round as “true” value, Y-axis = difference) along with a simple linear regression line of the difference (Y) on the average (X).

Regarding the validity of app results compared to NE (gold standard), only data from the second round were used and analyzed by scatter plot and modified Bland-Altman plot (X-axis = NE, Y-axis = difference) with a simple linear regression line. Simple linear regression line and its 95% confidence band is applied to determine if the difference between NE and each app result depends on the true value. Slope (b) of the linear regression line that is close to zero indicates that the difference between NE and each app result does not depend on the true value. P-value of slope that is greater than 0.05 reveals that the population slope is zero.

Paired t-test was performed to test the difference in duration of test (second) from the first and second round.

Statistical analyses were performed using PASW 18 and MedCalc Statistical software version 19.6.4.

RESULTS

There were 52 subjects (44 females) aged 21 to 54 with mean of 33.3 ± 7.9 years.

The mean logMar in each test, i.e., NE, AN, TE, and LC, were -0.06, -0.10, -0.08, and -0.04, respectively, in the first round of testing, and were -0.07, -0.12, -0.09, and -0.05, respectively, in the second round of testing (Table 1).

For test-retest reliability of NE and app results from the first and second round, ICC for NE and LC were quite high (0.712 and 0.789) compared to only 0.140 and 0.495 for AN and TE respectively (Table 2). The classic Bland-Altman plot of the difference between the first and second round revealed that the all mean differences were very close to zero with the mean (bias) of 0.01, 0.02, 0.02 and 0.01 for NE, AN, TE and LC respectively (Fig 3). However, NE has the narrowest 95% LOA of -0.04 to 0.07 whereas AN had the widest 95% LOA of -0.19 to 0.23. TE and LC had similar 95% LOA of -0.13 to 0.16 and -0.11 to 0.12 respectively.

<table>
<thead>
<tr>
<th>TABLE 1. Results of the tests from the first and second round.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>logMAR: Mean ± SD</strong></td>
</tr>
<tr>
<td><strong>First round</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Numeric EDTRS chart (NE)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>&quot;Chudjane test&quot; app</td>
</tr>
<tr>
<td>Numbers on app (AN)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Tumbling-E on app (TE)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Landolt-C on app (LC)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2. Test-Retest reliability between results from the first and second round.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICC</strong></td>
</tr>
<tr>
<td>Numeric EDTRS chart (NE)</td>
</tr>
<tr>
<td>&quot;Chudjane test&quot; app</td>
</tr>
<tr>
<td>Numbers on app (AN)</td>
</tr>
<tr>
<td>Tumbling-E on app (TE)</td>
</tr>
<tr>
<td>Landolt-C on app (LC)</td>
</tr>
</tbody>
</table>

* ICC (Intraclass Correlation Coefficient): 2-way random effect model, Absolute agreement, Single measure
Fig 3. Test-retest reliability for NE, AN, TE and LC.
The logMAR of each app test was compared to NE. The mean difference (bias) of NE - AN, NE - TE, and NE - LC were 0.04, 0.02 and -0.02 respectively, in the first round of testing, and 0.05, 0.02 and -0.03 in the second round of testing (Fig 4). The upper limit of all 95% CIs of the mean difference was not greater than 0.1 logMAR. Validity of app test were assessed using values from the second round via scatter plots, modified Bland-Altman plots and simple linear regression line (Fig 5). The 95% LOA for AN were a little wider than those for TE and LC (-0.11 to 0.20, -0.11 to 0.15 and -0.19 to 0.13 for AN, TE and LC respectively). The simple linear regression line of the difference between NE and each app result (Y) on NE (X) revealed that all slopes were close to zero and not statistically significant (slope = 0.222, 0.214 and 0.028; p-value = 0.402, 0.342 and 0.917 for AN, TE and LC respectively, (Table 3) indicating that the difference between NE and each app result did not depend on the NE.

The average duration (seconds) of the NE, AN, TE, and LC tests were 33.8, 56.8, 162.2 and 145.7, respectively, in the first round of tests, and 32.0, 49.4, 155.7, and 143.9, respectively, in the second round of tests (Table 4). The duration of the test from the first round was a little higher than the second round with the mean difference of 1.8, 7.4, 6.5 and 1.7 seconds respectively and not statistically significant (p-values > 0.1).

There were no adverse events reported from performing any of the tests.

**DISCUSSION**

The number of individuals who use smartphones and tablets is increasing annually. There has also been a significant increase in smartphone and tablets usage among health professionals. Currently, there are at least 100 vision test apps, but few have been validated. This study aimed to develop and validate a new visual acuity app for use on the iPhone and iPad that anyone could use to test their visual acuity by themselves.

This Chudjane app was a new design that aimed to decrease the gap in the optotype size, which is currently ten times of the standard EDTRS chart while still being based on logMAR scale progression to make the optotype size change more continuous. This would be expected to make the app more precise than tests performed using the old conventional chart. This app was also designed to decrease limitations of standard ETDRS chart that all fix characters were shown on light box chart. By using...
Fig 5. Validity of AN, TE and LC compared to NE (based on data from the second round).

TABLE 3. Difference between NE and app results from the first and second round.

<table>
<thead>
<tr>
<th></th>
<th>NE - AN</th>
<th>NE - TE</th>
<th>NE - LC</th>
</tr>
</thead>
<tbody>
<tr>
<td>First round</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean difference (Bias)</td>
<td>0.038</td>
<td>0.018</td>
<td>-0.021</td>
</tr>
<tr>
<td>SD</td>
<td>0.074</td>
<td>0.065</td>
<td>0.084</td>
</tr>
<tr>
<td>Second round</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean difference (Bias)</td>
<td>0.045</td>
<td>0.022</td>
<td>-0.026</td>
</tr>
<tr>
<td>SD</td>
<td>0.079</td>
<td>0.067</td>
<td>0.081</td>
</tr>
<tr>
<td>95% Limit of Agreement (LOA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean – 1.96 SD (Lower)</td>
<td>-0.109</td>
<td>-0.110</td>
<td>-0.186</td>
</tr>
<tr>
<td>Mean + 1.96 SD (Upper)</td>
<td>0.200</td>
<td>0.154</td>
<td>0.133</td>
</tr>
<tr>
<td>Linear regression of difference (Y) on NE (X)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>0.222</td>
<td>0.214</td>
<td>0.028</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(p=0.402)</td>
<td>(p=0.342)</td>
<td>(p=0.917)</td>
</tr>
</tbody>
</table>
TABLE 4. Duration of tests (second) in the first and second round.

<table>
<thead>
<tr>
<th>Optotype</th>
<th>Mean ± SD First round</th>
<th>Mean ± SD Second round</th>
<th>Difference (1st – 2nd round) Mean (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeric EDTRS chart (NE)</td>
<td>33.8 ± 14.3</td>
<td>32.0 ± 13.3</td>
<td>1.8 (-1.5 to 5.3)</td>
<td>0.28</td>
</tr>
<tr>
<td>“Chudjane test” App</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbers on app (AN)</td>
<td>56.8 ± 36.9</td>
<td>49.4 ± 21.8</td>
<td>7.4 (-1.7 to 16.5)</td>
<td>0.11</td>
</tr>
<tr>
<td>Tumbling-E on app (TE)</td>
<td>162.2 ± 85.2</td>
<td>155.7 ± 87.4</td>
<td>6.5 (-21.8 to 34.8)</td>
<td>0.65</td>
</tr>
<tr>
<td>Landolt ‘C’ on app (LC)</td>
<td>145.7 ± 83.6</td>
<td>143.9 ± 121.7</td>
<td>1.7 (-33.5 to 36.9)</td>
<td>0.92</td>
</tr>
</tbody>
</table>

randomized characters, multiple optotypes and more fine-scale in this app should decrease memorization effect and limitations of standard ETDRS chart.

In this study, the results revealed that the mean VA in the normal eye population was less than logMAR 0.00 or better than 6/6, which is the standard for normal eye VA.

The mean VA with AN was the best. This could be due to possibly three causes: First, there may have been a memorization effect taking place since this optotype on the app was not random and also more closely resembled the standard EDTRS chart. Second, the AN size did not continuously decrease the same way as for TE and LC on the app, and there was a larger step difference between lines for AN. Finally, AN involves greater recognition acuity (i.e., reading a letter or Arabic number is easier), which could make it easier to guess the answer than TE and LC optotypes with greater resolution acuity. Several studies have reported that recognition acuity is higher than resolution acuity in low-vision subjects and in normal-vision subjects.

In terms of the mean differences and agreement when comparing each optotype in the app to NE, the results were all satisfactory (less than 0.1 logMAR), although agreement with TE and LC was less than for AN, which is of practical significance that the margin of error between the two charts is less than one complete line, which would result in similar patient care decisions based on the visual acuity measurements from either chart. In terms of the mean difference of AN, it was the highest, which may result from a memorization effect occurring with AN, which was not the same as for the randomized optotypes, such as TE and LC. In a previous study, it was found that after a single VA test, a significant memory of a chart letter subset can be passively acquired and can persist for 10 days. The standard ETDRS back-illuminated light box had different brightness from the iPad digital screen display that use 100% brightness, also all the characters and lines were shown on the standard ETDRS light box but only a single character displayed on the iPad screen, these factors may cause memorization effect. The intraclass-Correlation-Coefficient (ICC) was highest in Landolt-C optotype because LC was optotype with greater resolution acuity, lesser memorization effect than NE, AN, TE and LC had smaller step difference between lines than AN in this app and NE.

The duration of the test using the app was longer than for the tests performed using the conventional ETDRS chart for all the optotypes, especially for TE and LC, since these were more detailed than for NE and AN by a factor of 10 times, and also, as many users were not acquainted with the use of the new app.

The strengths of this study include the fact that the test subjects tested themselves and that the LC and TE optotypes were random, which reduced subjective bias. There were three types of optotypes tested, so it is likely that the app can meet the needs of multiple populations and thus could be extensively applied.

The major limitation of this study to note is that the testing was only performed in a normal eye with best corrected visual acuity population. The use in clinical practice for an ophthalmic patient should also be studied in future research. The current design of the app interface on iPhone, which was used as a remote control, the AN optotypes that smaller than TE and LC may cause difficulty to use by people who have hyperopia and presbyopia. Still, this limitation could be eliminated by developing new voice to command features for the app. The app requires a little time for users to learn how to
use it. We will design it to be as user friendly as possible and further validation studies using other smart phones and screen type.

CONCLUSION

In conclusion, this study demonstrated that the “Chudjane” test application showed Landolt-C had highest test-retest reliability and validity compared to the standard ETDRS chart. This app is a new design self-assessment logMAR test for distance visual acuity suitable for normal eye population who have an iPhone and iPad as it allows the user to test and monitor their normal best corrected visual acuity by themselves. The use in clinical practice for an ophthalmic patient should also be studied in future research. Moreover, it has a potential to be a useful new tool to improve the services of physicians, such as screening and monitoring visual acuity.

ACKNOWLEDGMENTS

This research project was supported by Siriraj Research Fund, Grant number (IO) R016231009, Faculty of Medicine Siriraj Hospital Mahidol University and the authors gratefully acknowledge Dr. Chulaluk Komoltri from the Office for Research and Development for her assistance with statistical analysis.

Conflicts of interest: None

REFERENCES