Snellen and ETDRS Charts Compared Using a Computer Simulation

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>International Journal of Ophthalmology</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>IJO-2014-0310.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>03-Mar-2015</td>
</tr>
</tbody>
</table>
| Complete List of Authors:    | Shamir, Reuben; The Hebrew University, School of Engineering and Computer Science  
                              | Friedman, Yael; Hadassah – Hebrew University Medical Center, Ophthalmology  
                              | Joskowicz, Leo; The Hebrew University, School of Engineering and Computer Science  
                              | Mimouni, Michael; Rambam Medical Center, Ophthalmology  
                              | Blumenthal, Eytan; Rambam Medical Center, Ophthalmology |
| Keywords:                    | EDRS, Snellen, Computer Simulation, Visual Acuity Testing, Virtual Patients |
| Speciality:                  | others < Glaucoma, Others < Optometry & Ophthalmology, Blindness prevention and treatment |
Snellen and ETDRS Charts Compared Using a

Computer Simulation

Running Head: Comparing Snellen and ETDRS Charts

Reuben R. Shamir, PhD¹, Yael Friedman, MD², Leo Joskowicz, PhD¹, Michael
Mimouni, MD³, Eytan Z. Blumenthal, MD³

¹School of Engineering and Computer Science, The Hebrew University, Jerusalem
91120, Israel.

²Department of Ophthalmology, Hadassah – Hebrew University Medical Center,
Jerusalem 91120, Israel.

³Department of Ophthalmology, Rambam Medical Center, Haifa 31096, Israel.

Correspondence to: Eytan Z. Blumenthal, MD
Department of Ophthalmology, Rambam Medical Center
P.O.B 9602, Haifa, Israel 31096
Tel: 972-4-854-2668
Fax: 972-4-854-2142
E-mail: e_blumenthal@rambam.health.gov.il

Abbreviation of authors names:

Shamir RR, Friedman Y, Joskowicz L, Mimouni M, Blumenthal EZ

Conflict of Interest: None of the authors has conflict of interest with the submission.

Financial Support: No financial support was received for this submission.
Abstract

Aim: To compare accuracy, reproducibility and test duration for the Snellen and the Early Treatment of Diabetic Retinopathy Study (ETDRS) charts, two main tools used to measure visual acuity (VA).

Methods: A computer simulation was programmed to run multiple virtual patients, each with a unique set of assigned parameters, including visual acuity, false-positive and false-negative error values. For each virtual patient, assigned visual acuity was randomly chosen along a continuous scale spanning the range between 1.0 to 0.0 logMAR units (equivalent to 20/200 to 20/20). Each of 30,000 virtual patients were run ten times on each of the two VA charts.

Results: Average test duration (expressed as the total number of characters presented during the test ± S.D.) was 12.6 ± 11.1 and 31.2 ± 14.7 characters, for the Snellen and ETDRS, respectively. Accuracy, defined as the absolute difference (± S.D.) between the assigned VA and the measured VA, expressed in logMAR units, was superior in the ETDRS charts: 0.12 ± 0.14 and 0.08 ± 0.08, for the Snellen and ETDRS charts, respectively. Reproducibility, expressed as test retest variability, was superior in the ETDRS charts: 0.23 ± 0.17 and 0.11 ± 0.09 logMAR units, for the Snellen and ETDRS charts, respectively.

Conclusions: A comparison of true (assigned) VA to measured VA, demonstrated, on average, better accuracy and reproducibility of the ETDRS chart, but at the penalty of significantly longer test duration. These differences were most pronounced in the low visual acuity range. The reproducibility found using a simulation approach is in line with reproducibility values found in several clinical studies.

Key words: ETDRS; Snellen; Computer Simulation; Visual Acuity Testing; Virtual Patients
INTRODUCTION

Visual acuity (VA), an important measure reflecting the health of the eye, has been chosen as an endpoint in countless clinical studies, including trials on diabetic retinopathy,1, 2 macular degeneration,3 cataract surgery,4, 5 endophthalmitis,6 refractive surgery7 and others. Two commonly used tools for testing visual acuity in both the clinical and research setting include the Snellen and the Early Treatment of Diabetic Retinopathy Study (ETDRS) VA charts.8, 9

The Snellen VA chart, considered the most commonly used chart for measuring VA, consists of letters whose size is determined by the visual angle they occupy. The main drawbacks of this chart include the following: different rows have significantly different number of characters (Table 1), spanning from 1-8 characters per row; there is an inconsistent decrease in letter size from one row to the next; and not all presented characters are equally legible.8 The ETDRS chart was introduced in the 1980s (1), following recognition of the significant limitations of the Snellen chart, especially so in the low VA range. The main improvements incorporated into the ETDRS chart include: an equal number of characters per row (Table 1); an equal logarithmic decrement between successive rows; and the use of character types which are of relatively uniform legibility.8, 9 While it is generally agreed that the ETDRS chart has several distinct advantages over the Snellen chart, and has gained a role in clinical trials, its acceptance and penetration into the routine clinical encounter has been limited, perhaps related to the size and bulkiness of the chart, its cost, prolonged testing time and availability.

The reproducibility achieved during VA testing using each of these two commonly used charts has been previously studied.10-15 These clinical studies have focused on reproducibility, a common surrogate for accuracy, because a method for determining true VA, as opposed to measured VA, does not exist. For this study we created a computer simulation model to evaluate and compare the Snellen and ETDRS charts. Virtual patients were used to help quantify accuracy, reproducibility and test duration for testing with each of the above VA charts. Developing a simulation model provides us with the unique possibility to isolate and study the contribution of individual factors on accuracy in VA testing. It also has the advantage of allowing us to test VA accuracy directly, which we defined as the
absolute difference between the assigned VA and the measured VA, a comparison which would be impossible to measure by testing real patients.

Given the unique ability of computer simulations to study specific questions in a well defined and fixed setting, they have been used extensively in evaluating visual field (VF) algorithms\textsuperscript{16-19} as well as VF progression tools.\textsuperscript{20-22} In this study we exploit a computer simulation to provide data that can contribute towards developing a more refined and accurate test for VA.

**MATERIALS AND METHODS**

MATLAB, version 6.1 (The MathWorks, Natick, MA) a high level language computing and modeling software platform was used to create a computer simulation that tested the visual acuity of 30,000 virtual patients each carrying a unique set of parameters.

**Visual Acuity Charts**

The Revised 2000 Series ETDRS charts (Precision Vision, La Salle, Illinois) and the Traditional Snellen Eye Chart (Precision Vision, La Salle, Illinois) were modeled into the simulation.

**Visual Acuity Assignment**

Each of the virtual patients was randomly assigned a “true VA” from a continuous scale spanning the range 1.0 - 0.0 logMAR units, equivalent to 0.1- 1.0 decimal units, or 20/200 – 20/20 (Table 2). A frequency of seeing curve, analogous to the frequency of seeing curve described in relation to VF’s,\textsuperscript{23} was developed to describe a patient’s chance of correctly identifying a character of a given size, based on the patient’s true VA. The curve graphs the character size in logMAR units along the x-axis, and the probability (from 0-100%) of correctly determining the character on the y-axis (Figure 1). Based on a study that determined that changes of 0.2 logMAR or greater can be consistently differentiated from no change,\textsuperscript{24} the curve width was randomly assigned to span the range between 0.15 - 0.25 logMAR units with a uniform distribution.

**False positive and False Negative Responses**
In order to best simulate real VA testing, false positive and false negative responses were incorporated into the simulation. A false positive refers to a scenario where a virtual patient recognizes a character that based on the VA frequency of seeing curve, should not have been identified correctly. This reflects a real life situation where a real patient would correctly name a character from random guessing. With ten characters to choose from, the false positive rate modeled for this simulation was 10% on average (range: 0-20%).

A false negative response is an incorrect recognition of characters that according to the patients VA frequency of seeing curve should have been identified correctly. False negative errors reflect real life situations such as absent mindedness as well as errors in hearing and documenting on the part of the examiner. False negative responses were incorporated at an average rate of 5% (range: 0 - 10%).

**Test Termination Criteria**

Test termination criteria were established to avoid testing the virtual patient on each and every row, down to the bottom of the chart. In this study, the VA test terminates when $< 50\%$ of the characters in the row have been correctly identified, similar to the termination criteria used in most clinical settings. Of note, while VA is often tested in the clinic up until a VA of 20/20 (1.0, 6/6) and no further, both the Snellen and ETDRS charts contain character lines extending far beyond this cutoff. In this study we chose to recruit ‘virtual patients’ in the 6/60 - 6/6 VA range. However, each ‘virtual patient’ was tested as far down the chart as he/she could ‘see’, even beyond the 6/6 row, pending the ‘termination criteria’.

**Scoring Methods**

The algorithm by which a set of correct and incorrect character responses are translated into a ‘measured VA’ is known as the scoring method. In our study we chose the ‘smallest character row for which $\geq 50\%$ of the characters were correctly identified’, a common scoring method that would be equally feasible, and minimize bias, between the two charts.

Based on the unique set of assigned parameters including assigned VA, frequency of seeing curve width, false positive and false negative values, the simulation can compute for each character size presented, whether the patient would identify it
correctly. Each virtual patient underwent two separate VA tests, one using the Snellen and the other using the Original Series ETDRS chart. The testing order in a computer simulation is of no consequence. On each of the charts, the patient was presented every character of each row, one by one, starting from the top row and proceeding downwards, until the test termination criteria was fulfilled. Next, the scoring method was used to determine the measured VA.

**Reproducibility**

In previous clinical trials evaluating the Snellen and ETDRS VA charts, TRV has been used as a measure of reproducibility. Reproducibility values, expressed as test-retest variability (TRV), were determined by running each virtual patient through the VA simulation test ten times. TRV was calculated as ±1.96 standard deviations.

**Statistical analysis**

MATLAB, version 6.1 (The MathWorks, Natick, MA), and JMP statistical software version 5.0 (SAS Institute, Cary, NC) were used to analyze the data. Statistical analysis performed in this study included descriptive statistics and the paired t-test used for comparison of the Snellen and ETDRS chart results. P-values <0.05 were considered statistically significant.

**RESULTS**

Table 3 provides accuracy, reproducibility and test duration data, separately, for both the Snellen and ETDRS charts. In addition, each of these values, across the two charts, is compared for statistical significance. This analysis is provided for the group as a whole, as well as for each of the three VA sub-groups: high, moderate and low VA.

The most noticeable difference between the two charts relates to the length of the VA examination, as measured by the average number of characters presented during the test. The test duration, for Snellen and ETDRS charts, respectively, was found to be: 12.6 and 31.2 for the entire group, 25.3 and 47.7 for high VA subjects, 9.0 and 31.4 for moderate VA subjects and 3.6 and 14.9 for low VA subjects. Reproducibility, as well as accuracy, both for the group as a whole, as well as in the sub-group analysis, was better for the ETDRS chart, as listed in Table 3. In the population sub-group analysis presented in Table 3, the 30,000 ‘virtual patients’ were sub-divided into three
Subgroups based on their assigned VA. This sub-group analysis was performed to determine whether the overall results found in this study equally pertain to patients with high, moderate and low VA.

In respect to accuracy, for the Snellen chart, the low VA group demonstrated, on average, a somewhat larger error, while, for the ETDRS chart, the inverse was found, with the best accuracy achieved for the low VA group. In respect to reproducibility, both charts demonstrated the tightest reproducibility for the low VA group.

Table 3 highlights the differences found between the two charts. The ETDRS chart was found to be, overall, more accurate and reproducible in both the analysis and sub-analysis comparisons. However, with a penalty of significantly longer test duration, found on average to be more than doubled.

DISCUSSION

Many studies evaluating diagnostic tools report reproducibility, rather than accuracy, since accuracy cannot be evaluated in a scenario lacking a superior gold standard, which the outcome of the test could be compared against. Hence, in this simulation study we receive a relatively rare opportunity to evaluate how reproducibility and accuracy might fare with each other. Table 3 demonstrates that mean accuracy was found to be roughly half the reproducibility, when reproducibility is presented as TRV.

It is of interest to compare the reproducibility found using our simulation approach against studies performed on real patients. Our current study, on ‘virtual patients’, found an overall TRV of 0.23 and 0.11 logMAR units, for Snellen and ETDRS charts, respectively. In comparison, prior studies performed on real subjects have found TRV spanning 0.24 - 0.33 logMAR for Snellen and 0.15 - 0.20 logMAR for ETDRS charts.14,25,26

The ETDRS chart is often considered superior to the Snellen chart. While our data strengthen this conclusion, we highlight significant discrepancies in test duration, which might, as the sole differentiating factor, allow the longer test to simply be proportionally more accurate. This finding is supported by previous studies that have shown that the time taken to complete the ETDRS is longer than that taken to complete the Snellen chart.27 Test (or actually number of character) duration differences relates to the difference in the number of characters on each chart in the 6/60 to 6/6 range, specifically 36 in the Snellen chart and 55 in the ETDRS chart.
addition, the longer duration may be related to the different design of the VA chart as there are less characters in the top rows of the Snellen chart and the examination was performed from the top row and proceeded downwards. It remains to be determined whether the ETDRS chart remains superior even after adjustments are made for test duration.

A simulation approach can analyze aspects of the chart layouts which cannot be studied on ‘real patients’, and, in particular, address the concept of ‘accuracy’. Nevertheless, the approach utilized in this study, harbors several limitations common to all simulation studies, primarily, that the human response cannot be reliably modeled in full. Of note, we did not address the known differential readability of different characters comprising the two character sets. We did not model a ‘learning effect’, nor a ‘tiring effect’, and, perhaps most illusive, a simulation approach is inherently unable to model unexpected deviations from the expected human response, which in the context of VA is expected to follow the ‘S-shaped’ frequency of seeing curve. We measured exchanged test duration as for the number of letters presented rather than overall time needed to record visual acuity, neither of which is a validated method of measuring visual acuity testing duration because tests were virtually run in a computer simulation setting. As we chose to recruit ‘virtual patients’ in the 6/60 - 6/6 VA range, this study’s conclusions cannot be applied to those with a visual acuity outside of this range. As each character is presented to the virtual patient one by one, this study did not model the crowding phenomenon encountered in real life charts.

The testing of VA, which provides the “vital signs” of an eye exam, is a relatively neglected field. A computer simulation approach may assist in evaluating strengths and weaknesses of various chart designs and testing algorithms, in the hope of improving the chart layout, and ultimately devising more sophisticated and accurate computerized methods for measuring VA in both the clinical and research setting.

ACKNOWLEDGEMENTS

Conflicts of Interest: None of the authors has conflict of interest with the submission.

Financial Support: No financial support was received for this submission.
REFERENCES


19. Turpin A, McKendrick AM, Johnson CA, Vingrys AJ. Properties of perimetric threshold estimates from full threshold, ZEST, and SITA-like strategies, as


FIGURE LEGENDS

Figure 1: Probability of correctly determining the character of a certain LogMAR unit.
### Table 1. A description of each of the 8 Snellen and 11 ETDRS chart rows, spanning the VA range: 6/60 - 6/6.

<table>
<thead>
<tr>
<th>Line (decimal)</th>
<th>Visual Acuity (decimal)</th>
<th>Metric/Feet equivalent</th>
<th>Number of characters</th>
<th>Visual Acuity (logMAR equivalent)</th>
<th>Visual Acuity (decimal)</th>
<th>Metric/Feet equivalent</th>
<th>Number of characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>6/60; 20/200</td>
<td>1</td>
<td>0.1</td>
<td>1.0</td>
<td>6/60; 20/200</td>
<td>5</td>
</tr>
<tr>
<td>0.2</td>
<td>0.7</td>
<td>6/30; 20/100</td>
<td>2</td>
<td>0.125</td>
<td>0.9</td>
<td>6/48; 20/160</td>
<td>5</td>
</tr>
<tr>
<td>0.3</td>
<td>0.52</td>
<td>6/20; 20/70</td>
<td>3</td>
<td>0.16</td>
<td>0.8</td>
<td>6/38; 20/125</td>
<td>5</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>6/15; 20/50</td>
<td>4</td>
<td>0.20</td>
<td>0.7</td>
<td>6/30; 20/100</td>
<td>5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3</td>
<td>6/12; 20/40</td>
<td>5</td>
<td>0.25</td>
<td>0.6</td>
<td>6/24; 20/80</td>
<td>5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.22</td>
<td>6/10; 20/30</td>
<td>6</td>
<td>0.32</td>
<td>0.5</td>
<td>6/19; 20/63</td>
<td>5</td>
</tr>
<tr>
<td>0.8</td>
<td>0.1</td>
<td>6/7.5; 20/25</td>
<td>7</td>
<td>0.40</td>
<td>0.4</td>
<td>6/15; 20/50</td>
<td>5</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0</td>
<td>6/6; 20/20</td>
<td>8</td>
<td>0.50</td>
<td>0.3</td>
<td>6/12; 20/40</td>
<td>5</td>
</tr>
</tbody>
</table>
Shamir et al
Snellen and ETDRS compared using a computer simulation

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.63</td>
<td>0.2</td>
<td>6/9.5; 20/32</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>0.1</td>
<td>6/7.5; 20/25</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.0</td>
<td>6/6; 20/20</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. ‘Virtual patient’ characteristics: ‘intended values’ compared to ‘assigned values’ randomly generated for each of the 30,000 ‘virtual patients’.

<table>
<thead>
<tr>
<th></th>
<th>n (virtual patients)</th>
<th>VA (logMAR units)</th>
<th>Width of ‘frequency of seeing curve’ (logMAR units)</th>
<th>False negative rates (%)</th>
<th>False positive rates (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All VA (0.0-1.0)</td>
<td>Intended (mean):</td>
<td>30,000</td>
<td>0.5</td>
<td>0.2 (range: 0.15 - 0.25)</td>
<td>5 (range: 0-10)</td>
</tr>
<tr>
<td></td>
<td>Assigned (mean ± SD):</td>
<td>30,000</td>
<td>0.502 ± 0.289</td>
<td>0.2002 ± 0.0289</td>
<td>4.99 ± 2.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High VA (0.0-0.33)</td>
<td>Intended (mean):</td>
<td>10,000</td>
<td>0.165</td>
<td>0.2 (range: 0.15 - 0.25)</td>
<td>5 (range: 0-10)</td>
</tr>
<tr>
<td></td>
<td>Assigned (mean ± SD):</td>
<td>9,863</td>
<td>0.165 ± 0.095</td>
<td>0.2002 ± 0.0289</td>
<td>4.94 ± 2.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate VA (0.33 - 0.67)</td>
<td>Intended (mean):</td>
<td>10,000</td>
<td>0.5</td>
<td>0.2 (range: 0.15 - 0.25)</td>
<td>5 (range: 0-10)</td>
</tr>
<tr>
<td></td>
<td>Assigned (mean ± SD):</td>
<td>10,087</td>
<td>0.500 ± 0.097</td>
<td>0.2003 ± 0.0288</td>
<td>5.01 ± 2.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low VA</td>
<td>Intended (mean ± SD):</td>
<td>10,000</td>
<td>0.835</td>
<td>0.2 (range: 0.15 - 0.25)</td>
<td>5 (range: 0-10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Shamir et al
Snellen and ETDRS compared using a computer simulation

<table>
<thead>
<tr>
<th></th>
<th>Assigned (mean ± SD)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10,050</td>
<td>0.834 ± 0.095</td>
<td>0.25</td>
<td>5.02 ± 2.90</td>
</tr>
<tr>
<td></td>
<td>0.2001 ± 0.0289</td>
<td>9.99 ± 5.75</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

VA: visual acuity, SD: standard deviation.
**Table 3. A comparison of the Snellen and ETDRS chart results**

<table>
<thead>
<tr>
<th>Population subgroups</th>
<th>Snellen Accuracy (TRV)</th>
<th>Reproducibility</th>
<th>Duration</th>
<th>ETDRS Accuracy (TRV)</th>
<th>Reproducibility</th>
<th>Duration</th>
<th>Statistical comparison¹: Snellen vs. ETDRS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Snellen vs. ETDRS</td>
</tr>
<tr>
<td>All VA (0.0-1.0)</td>
<td>0.119 ± 0.139</td>
<td>0.232 ± 0.167</td>
<td>12.6 ±</td>
<td>0.077 ± 0.076</td>
<td>0.107 ± 0.085</td>
<td>31.2 ±</td>
<td>&lt; .0001 &lt; .00001 &lt; .00001</td>
</tr>
<tr>
<td>High VA (0.0-0.33)</td>
<td>0.119 ± 0.185</td>
<td>0.277 ± 0.227</td>
<td>25.3 ±</td>
<td>0.088 ± 0.088</td>
<td>0.134 ± 0.114</td>
<td>47.7 ±</td>
<td>&lt; .0001 &lt; .00001 &lt; .00001</td>
</tr>
<tr>
<td>Moderate VA (0.33 - 0.67)</td>
<td>0.117 ± 0.185</td>
<td>0.216 ± 0.227</td>
<td>9.0 ±</td>
<td>0.078 ± 0.076</td>
<td>0.107 ± 0.069</td>
<td>31.4 ±</td>
<td>&lt; .0001 &lt; .00001 &lt; .00001</td>
</tr>
<tr>
<td>Low VA (0.67 - 1.0)</td>
<td>0.122 ± 0.127</td>
<td>0.205 ± 0.125</td>
<td>3.6 ±</td>
<td>0.065 ± 0.069</td>
<td>0.080 ± 0.052</td>
<td>14.9 ±</td>
<td>&lt; .0001 &lt; .00001 &lt; .00001</td>
</tr>
</tbody>
</table>

¹Paired t-test

TRV: Test retest variability
Probability of correctly determining the character of a certain LogMAR unit.
24x15mm (600 x 600 DPI)