VALIDITY AND REPEATABILITY OF COMPUTERIZED VISUAL ACUITY DEVICES, SS II AND EVA, FOR HOTV VISUAL ACUITY MEASUREMENTS IN YOUNG HEALTHY ADULTS UNDER HABITUAL-CORRECTED AND MYOPIC-BLUR CONDITIONS

A thesis presented to the graduate faculty of
The New England College of Optometry in partial fulfillment
of the requirements for the degree of Master of Science

Elena A Zaharova, OD

February, 2011

©Elena A Zaharova
All rights reserved

The author hereby grants the New England College of Optometry permission to reproduce and to distribute publicly paper and electronic copies of the thesis document in whole or in part.

VALIDITY AND REPEATABILITY OF COMPUTERIZED VISUAL ACUITY DEVICES, SS II AND EVA, FOR HOTV VISUAL ACUITY MEASUREMENTS IN YOUNG HEALTHY ADULTS UNDER HABITUAL-CORRECTED AND MYOPIC-BLUR CONDITIONS

Elena A. Zaharova, OD

This manuscript has been read and accepted by the
Thesis Committee in satisfaction of the
thesis requirement for the degree of Master of Science

2/11/11
Date

Bruce Moore, OD
Graduate Faculty Advisor

Luisa Mayer, MEd, PhD

Frank Thorn, PhD, OD

Julie A. Blacksmith, OD

4-28-11
Date

Nancy J. Coletta, OD, PhD
Director of Graduate Studies

The New England College of Optometry

VALIDITY AND REPEATABILITY OF COMPUTERIZED VISUAL ACUITY DEVICES, SS II AND EVA, FOR HOTV VISUAL ACUITY MEASUREMENTS IN YOUNG HEALTHY ADULTS UNDER HABITUAL-CORRECTED AND MYOPIC-BLUR CONDITIONS

Elena A Zaharova, OD

The New England College of Optometry, 2011

Abstract:

Purpose: To evaluate two computerized visual acuity testing devices, Smart System II PC (SS II) by M&S Technologies, and Electronic Visual Acuity Tester (EVA) by Haag-Streit USA, in terms of accuracy and test-retest repeatability as compared to a standard back-illuminated logMar HOTV chart under habitual-corrected and myopic blur-induced conditions.

Methods: Fully-interpolated 5 letters per line HOTV monocular visual acuities of the right eye were obtained from 30 young healthy adults on two visits separated by 2 weeks. The stimulus was a single standard letter optotype (H, O, T, and V) surrounded by crowding bars. The tests were presented in a randomized order by one examiner. The acuity test results were compared to results from the logMAR HOTV chart.

Results: *SS II repeatability*: No statistically significant difference was found between VA measurements between week 1 and week 2 ($p>0.05$) under either habitually-corrected or blur-induced conditions. No statistically significant differences were found when comparing the 95% limits of agreement of repeatability of visual acuity results for SSII vs. HOTV chart ($p > 0.20$) for either condition. *EVA repeatability*: No statistically significant differences

were found between VA measurements between week 1 and 2 ($p>0.05$) under either habitually-corrected or blur-induced conditions. The limits of agreement were significantly larger for HOTV chart acuities compared to EVA acuities ($p < 0.05$) for both conditions. *SS II Validity*: no significant difference visual acuity results were found between SS II test and HOTV chart for either the habitually-corrected condition ($p = 0.12$) or blur-induced condition ($p=0.90$). Limits of agreement between SSII and HOTV chart ranged from +0.104 to -0.141 logMAR in the habitually-corrected condition and from +0.353 to -0.345 logMAR under the blur-induced condition. *EVA Validity*: Under habitually-corrected conditions EVA underestimated visual acuity measurements by a mean -0.087 logMAR ($p<0.05$). No significant difference in visual acuity estimation was found between EVA test and logMAR HOTV chart under the blur-induced condition ($p=0.89$). The limits of agreement between the two instruments ranged from +0.056 to -0.230 logMAR in the habitually-corrected condition and from +0.246 to -0.253 logMAR under the blur-induced condition.

Conclusions: The commercial computerized visual acuity testers, SS II and EVA, were easy to administer to adult inexperienced participants. These computerized systems can measure VA as reliably as the logMAR HOTV chart, thus providing viable alternatives to the crowded HOTV visual acuity measurements in young healthy adults. Larger variability of VA measurements should be expected under myopic blur conditions than habitual-corrected conditions for all measurement systems tested.

Acknowledgements Supported by NEI grants R24 EY014817 and T35 007149

None of the authors had financial interest in any of the devices

Table of Contents

General Introduction Page 1
Methods Page 9
Results Page 25
Discussion and Conclusions Page 42
Bibliography Page 54

List of Figures and tables

Figure 1: The appearance of a single crowded optotypes, page 4
Figure 2: logMAR HOTV chart, page 11
Figure 3: Back-illuminated part of logMAR HOTV chart, page 11
Figure 4: Smart System II PC (SS II), page 13
Figure 5: EVA, page 14
Figure 6: Differences in visual acuity estimation comparing Interpolation of Pass/Fail (Interpol LogMar) and Pass/Fail (P/F) threshold method under habitually-corrected conditions, page 28
Figure 7: Differences in visual acuity estimation comparing Interpolation of Pass/Fail (Interpol LogMar) and Pass/Fail (P/F) threshold method under myopic blur-induced conditions, page 29
Figure 8: Bland-Altman plot of VA measurements for HOTV vs. SSII under habitual corrected conditions, page 36
Figure 9: Bland-Altman plot of VA measurements for HOTV vs. SSII under blur-induced conditions, page 37
Figure 10: Bland-Altman plot of VA measurements for HOTV vs. EVA under habitual corrected conditions, page 40
Figure 11: Bland-Altman plot of VA measurements for HOTV vs. EVA under blur-induced conditions, page 41

Table 1: Amblyopia Treatment Study Visual Acuity Testing Protocol, page 17-18
Table 2. Summary of Analyses, page 26
Table 3: Repeatability of VA measurements for HOTV and SSII, page 31
Table 4: Repeatability of VA measurements for HOTV and EVA, page 33
Table 5: Validity of SSII VA measurements as compared to HOTV, page 35
Table 6: Validity of EVA VA measurements as compared to HOTV, page 39
Table 7. Summary and Interpretation of Findings, page 43
Table 8: Previous studies showing repeatability of visual acuity measurements, page 47-48

GENERAL INTRODUCTION

Visual acuity (VA) assessment is a critical component of every eye examination. If performed correctly, a visual acuity measurement will help identify individuals with significant vision problems. This is particularly important in children, where high refractive error, amblyopia, strabismus, and ocular disease may permanently impair vision.

The retroilluminated plastic ETDRS visual acuity chart was developed for adult vision clinical research, and it is still widely used in clinical research studies (Bailey and Lovie, 1976; Ferris and Bailey, 1996; NRC, 2002). In clinical trials, when precise quantitative assessment of visual acuity is required, standardization of VA measurements is particularly important. The design of ETDRS chart is based on the chart created by Bailey and Lovie (Bailey and Lovie, 1976), which incorporated recommendations from the National Research Council (NRC) (NRC, 2002). The most basic principle of the chart's design is that the visual acuity task for all letter sizes is practically the same and remains the same when nonstandard viewing distances are used. The Sloan letters that are used in the ETDRS chart have similar difficulty scores when compared to each other and to a Landolt Ring (Bailey and Lovie, 1976; Ferris and Bailey, 1996). The spacing between the letters is equal to one-letter width (4 stroke widths) and the space between lines is equal in height to the letters of the next lower line (Bailey and Lovie, 1976; Ferris and Bailey, 1996). The progression of letter sizes follows a geometric progression whose multiplier is equal to the tenth root of ten or 0.1 log unit (1.2589) (Bailey and Lovie, 1976; Ferris and Bailey, 1996), thus keeping the scaling factor constant throughout the chart and unaltered when nonstandard distances are used. This chart characteristic (logMAR spacing between optotype lines) has become an

accepted standard in VA measurement for research study as well as clinical acuity testing. The ETDRS chart, first introduced by Ferris (Ferris at al, 1992), conforms with the guidelines for standardized visual acuity measurements described by Ferris and Bailey (Bailey and Lovie, 1976; Ferris and Bailey, 1996), the only exception being the use of 5 optotypes at each size level instead of recommended 10 by NAS. Even though the reduced number of optotypes could theoretically increase the probability of false recognition due to guessing, it does not seem to affect the final result of the measurement (ETDRS, 1980). Retroillumination of the ETDRS chart allows for long-term stable lighting with high contrast. To ensure accurate and reproducible assessment of visual acuity, it is therefore recommended to use the ETDRS chart for clinical research.

In contrast to visual acuity testing designed for adults, isolated HOTV optotypes, with and without crowding bars, are widely used for testing young children. Visual acuity assessment requires a certain cognitive level, causing practical difficulties when testing preschool children. The ETDRS chart had been successfully used in children at 5½ years (Cryotherapy for Retinopathy of Prematurity Cooperative Group, 1996; Rice et al, 2004). However, it is difficult for pre-literate children to complete visual acuity testing with the ETDRS letter chart, because of their unfamiliarity with letters and the complexity of the task. Therefore, an important consideration for testing children in the 3 to 5 year age group is the selection of a test that is appropriate to the child's age. Children's ability to complete visual acuity testing increases with age (Merritt, 1996; VIP, 2004b). Several different picture and optotypes charts that are easy for a child to understand and complete have been developed. Allen picture acuity test has high testability in pre-literate children but it overestimates visual

acuity in amblyopia (De Young-Smith and Baker, 1986; Mayer and Gross, 1990) and, therefore, is not used widely. The HOTV letter-matching test, which is the simplest form of The Screening Test for Young Children and Retardates (STYCAR) (Sheridan, 1960) was first described by Lippmann (Lippmann, 1969) and extensively studied by Friendly (Friendly, 1978) in 1978. The HOTV test and Lea Symbols, a symbol-matching test, have won the most recognition over the past years due to high testability rates and success in amblyopia screening (De Young-Smith and Baker, 1986; Hered, 1997; Merritt et al, 1996, VIP, 2003; VIP, 2004b). HOTV testability for 3-year-old children had been reported to be between 64% and 96%, whereas testability for the Lea Symbols ranged from 75% to 97.2%, making the HOTV test to be a slightly more difficult test than Lea symbols for 3-year-old children (Hered, 1997; Merritt et al, 1996, VIP, 2003; VIP, 2004b). No significant difference in testability or between-test agreement was found for children ages 4 to 5 years old. Both tests showed high between-test agreement, as much as 93%-100% for HOTV visual acuity testing and 97%-99.9% for Lea symbol visual acuity testing (Hered, 1997; Merritt et al, 1996, VIP, 2004). HOTV optotypes had been reported to provide better visual acuity results with the smallest 8.6% of missed diagnosis of amblyopia when compared to the Snellen, E test and Lea Symbols (De Young-Smith and Baker, 1986).

For effective detection of amblyopia in children, vision screening has been recommended between 3 and 5 years of age. To further increase testability for preschool vision screening, presentation of a single optotype rather than lines of optotypes have been widely used. It was reported that the presentation of a single optotype makes it easier for preschool children to identify the presented letter (Friendly, 1978). Crowding bars

surrounding the isolated optotype have been implemented to create a test that is more comparable to the crowding of optotypes on a line, and thus potentially improve the detection of amblyopia (Flom, 1963; Muller, 1951; Stuart and Burian, 1962). Muller in 1951 found that increasing the spacing between letters resulted in improved visual acuity in amblyopia. Flom (Flom, 1963) further studied spatial aspects of interaction between peripheral contours by varying the separation of crowding bars and the isolated letter. He found that the detection of the gap in amblyopic eyes was impaired by the crowding bars when they were displaced by as much as 14 to 23 min arc, as compared to 4 min arc in normal eyes, showing that extent of the zone of interaction is angularly larger for amblyopic eyes. In comparison to testing with single letters alone, crowded isolated optotypes are thought to improve detection of amblyopia (Flom, 1963; Stuart and Burian, 1962). In this study isolated optotypes surrounded by crowding bars of the same stroke width at a distance of half the optotype's height from the optotypes will be used (Figure 1).

Figure 1: The appearance of a single crowded optotype



Advances in technology have allowed computerized versions of visual acuity stimuli to be widely accepted in optometric and ophthalmologic practices. The use of computerized visual acuity charts offers many advantages such as standardization of screen luminance and letter contrast, random presentation of optotypes (either isolated or in single / multiple lines) and logarithmic progression across the entire range of acuity levels. Computerized visual acuity test had been shown to provide fast, reliable and accurate measurements in adults and children (Holmes et al, 2001; McClenaghan et al, 2007; VIP, 2004a). Currently there are many computerized versions of visual acuity tests available: ProVideo (Innova systems, inc), Acuity Pro (VisionScience Software), Lombart CVSi21 (Lombart Intrument Co) and many others. Typically these charts offer a wide selection of different letter-based or symbol-based visual acuity charts with variety of testing formats under remote control. Many of these systems are available using either PC or Apple Macintosh computers and appropriate software. Some systems are in portable versions. Two different computerized visual acuity devices that utilize HOTV optotypes with crowding bars will be investigated in this study. These are the Smart System II PC (SS II) by M&S Technologies and Electronic Visual Acuity (EVA) test by Haag-Streit USA.

The Smart System II PC (SS II) by M&S Technologies is one of the most widely distributed computerized visual acuity testers and is currently being used by multiple well-known institutions such as Vistakon, The Vision Care Institute, LLC. The main advantage of the system is that it allows the examiner a high degree of control over the type of visual acuity test performed. SS II can be calibrated for lane distances of 6 to 22 feet, which allows performing the test at non-standardized distances without an effect on visual acuity

measurements (Bailey and Lovie, 1976). The system has an advantage of generating a random order of optotypes that can be presented as a single optotype, isolated crowded optotypes or as a line of optotypes. A recent model of SS II is available as PC software using a PC computer which allows for greater portability. Studies have shown that SS II can provide an accurate alternative to the ETDRS chart test of visual acuity in adults (McClenaghan et al, 2007). Even though the visual acuity measurement procedure is simple to administer, SS II has not been yet tested in a pediatric population.

EVA has been widely used in multiple multicenter clinical trials, including the NIH funded research projects Amblyopia Treatment Studies (ATS) (Cotter et al; 2007; Rice et al, 2004) and Vision in Preschoolers (VIP) Study (Kulp et al, 2004), by testing visual acuity of preschool children with isolated, crowded HOTV optotypes. The main advantage of the EVA system is the fact that it uses an automated test procedure for testing children developed by the researchers in the Amblyopia Treatment Study (ATS) for use with the Baylor Video Acuity Tester. Automation ensures strict adherence to the testing procedure, thus minimizing the variability related to the tester. This protocol is not difficult to teach, learn, and implement (Holmes et al, 2001). Studies had shown high testability for EVA visual acuity testing in preschool children: 93.3% of 3.5 year-olds and 94%-98.8% of 5 year olds were able to complete monocular visual acuity testing for crowded isolated HOTV optotypes (Holmes et al, 2001; VIP, 2004b).

Since many versions of computerized visual acuity tests are widely used for screening purposes when patients have uncorrected refractive errors, it is important to evaluate the performance of computerized visual acuity testers under blur-induced conditions. Visual

acuity is a sensitive measure that may differ based on differences in contrast and illumination levels (and differences in optotype). Blur has been shown to decrease contrast sensitivity, more so for high than for low spatial frequencies (Campbell and Green, 1965). Previous studies have shown that differences in background luminance, stimulus contrast, and dioptric blur can affect visual acuity measurements (Johnson and Casson, 1995). Moreover, it had been shown that the effect of these factors is essentially additive (Johnson and Casson, 1995), making it imperative to consider them when determining the threshold visual acuity with computerized versions of VA measurement. One diopter blur is the amount of optical blur that shows significant sensitivity loss for the stimuli with high spatial frequencies (Thorn, 1990), and, therefore, is directly related to the visual acuity. Optical blur induced with +1.25 DS lens was found to reduce visual acuity at all luminances tested in one study (Simpson et al, 1986). It is slightly less than the mean amount of blur introduced by accommodation to intermediate focus in darkness. Induced-myopic blur with +1.25 DS lens which will be used in the study will allow us to assess differences between the stimuli when measuring subjects with suboptimal vision.

There have been a limited number of published studies to support the validity of computerized HOTV systems compared to standard logMAR format charts. Rice et al (Rice et al, 2004) compared EVA with an electronic-ETDRS chart in children aged 5 to 12 years and found that visual acuity performance was better with EVA than the E-ETDRS chart with a median difference of 0.06 log MAR. McClenaghan et al (McClenaghan et al, 2007) evaluated electronic-ETDRS presented by the SS II system against the gold standard ETDRS chart with comparable results between the two charts. However, to date there is no clinical

evidence to support the validity of computerized HOTV systems when compared to the gold standard ETDRS visual acuity chart. In preschool vision screening or in multicenter clinical trials standardization of visual acuity testing is essential to reduce uncontrolled variability across clinical trials. It is important to determine how well computerized visual acuity charts compare to the gold standard to ensure accurate and reproducible assessment of visual acuity measurements which is an essential component for managing pediatric eye disease. The aim of this work is to determine how accurately the SSII and EVA tests measured HOTV recognition acuity compared to a standard logMAR format chart using HOTV optotypes. To evaluate accuracy, this study will investigate inter-subject variation and test repeatability through multiple measurements of visual acuity under habitual-corrected and blurred conditions.

METHODS

I. Participants

Thirty optometry students were recruited to participate in the study; subjects were required to meet all of the following criteria: 1) emmetropia or moderate refractive error, between +2.50 and -6.00 D sphere with 2.0 D or less of astigmatism, and anisometropia of 1.5 D or less, as determined from measuring their habitual refractive correction; 2) habitually-corrected visual acuity of 20/20 or better in the right eye, verified at the beginning of each session by ETDRS Chart; 3) no history of amblyopia, strabismus, ocular surgery, or current or previous eye disease other than resolved anterior segment inflammatory/infectious disease. The same optical correction, whether glasses or contact lenses, was required to be worn for both test sessions. The number of subjects for the present study was determined using an *a priori* power calculation, in order to identify differences between visual acuity of 0.1 logMAR or greater with statistical power > 80%. The criterion difference of 0.1 logMAR was chosen on the basis of a literature review of clinical relevance and inter-test reliability.

The average age of participants was 24.6 years old with a standard deviation of 22.3 to 26.9 years. There were 19 females and 11 males. Mean spherical equivalent refractive error of the right eye was -1.84D (SD +/- 1.79D). There were 10 participants who did not wear any correction, 10 participants who wore glasses and 10 who wore contact lenses. The mean habitually-corrected visual acuity measured with the ETDRS chart was -0.11 logMAR (SD +/- 0.06 logMAR) at week 1 and -0.12 logMAR with a standard deviation of +/-0.08 logMAR at week 2 test (p=0.45).

All participants gave written informed consent to participate in the study. Study protocol and all study procedures were approved by the Institutional Review Board of the New England College of Optometry and met International Standards for Human Medical Experimentation (WMA Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects, http://www.wma.net/en/30publications/10policies /b3/index.html).

II. Apparatus Description

1. Standard HOTV chart (logMAR HOTV)

Even though the ETDRS Sloan letter chart is widely considered to be the standard for clinical visual acuity research, in this study a modified version of the ETDRS letter chart using only four Sloan letters (H, O, T, V) was used as the standard visual acuity test because these are the four optotypes presented by SS II and EVA. The logMAR format HOTV chart (Good-Lite, Co. Elgin, IL) (see Figure 2) is a back-illuminated printed plastic chart. It has LED lighting system that illuminates the chart from the back and results in even illumination over the entire visual acuity chart surface (see Figure 3). Sloan letters H, O, T, and V, are presented in lines with a 0.1 logMAR progression between lines as recommended by NAS (Bailey and Lovie, 1976; Ferris and Sperduto, 1982; NRC, 2002). Five optotypes are printed on each line, with the exception of the largest line on which there are four optotypes. Each optotype has an aspect ratio of height to width of 1:1 and stroke width 1/5 of height/width. Each optotype is surrounded by four bars, unconnected at the four corners, with the same stroke width of 1/5 height/width of optotype and an angular separation of ½ height/width

from the outer edge of the optotype. The crowded optotypes are isolated from each other by 2 times the height/width of the optotype to ensure that the crowding effect comes from the crowding bars only. There are 11 lines on the full chart, spanning acuity from +0.70 logMAR to -0.30 logMAR. Mean luminance of white background was measured as 30.5 ft-cd (105 cd/m^2) and Michelson contrast of letters against background was measured as 0.936 (Weber contrast 0.967) by a Spectra Pritchard Photometer at a distance of 1.5 meters with an aperture size of 6 minarc.

Figure 2: logMAR HOTV chart

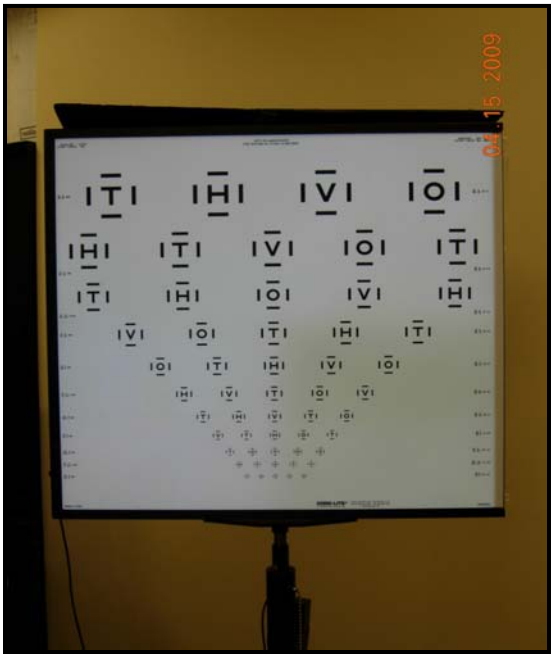


Figure 3: back-illuminated part of logMAR HOTV chart



2. Comparison device: The Smart System II PC (SS II)

SS II (M&S technologies, Inc. Skokie, IL USA) is a digital system that consists of a 17-inch flat LCD monitor with built-in PC-based visual acuity software and a remote controller (see Figure 4). The optotype stimuli follow Snellen visual acuity progression with approximate intervals of 0.1 logMAR as recommended by NAS (Bailey and Lovie, 1976; Ferris and Sperduto, 1982; NRC, 2002). H, O, T or V optotypes are presented one at a time with a total of 5 optotypes presented at each logMAR visual acuity level. Each optotype has an aspect ratio of height/width of 1:1 and stroke width 1/5 of height/width. Each optotype is surrounded by crowding bars with same stroke width of 1/5 height/width and an angular separation of ½ height/width from the optotype. SS II was programmed to generate random

sequences of non-repeatable back to back optotypes to ensure similar variability as the standard logMAR HOTV printed chart. The mean luminance of the LCD monitor background was 27.3 ft-cd (97.5 cd/m^2) and Michelson contrast was 0.943 (Weber contrast 0.970) measured by a Spectra Pritchard Photometer at a distance of 1.5 meters with aperture size of 6 minarc.

Figure 4: Smart System II PC (SS II)



3. Comparison device: EVA

EVA (Haag-Streit USA, Mason, OH) is a computerized device that includes a 17-inch CRT type monitor (Baylor-Video Acuity Tester) with built-in ATS Software and a handheld Palm Pilot (see Figure 5). The optotype stimuli are in intervals of 0.1 logMAR as recommended by NAS (Bailey and Lovie, 1976; Ferris and Sperduto, 1982; NRC, 2002) with

the H, O, T or V optotype presented one at a time at each logMAR visual acuity level. Each optotype has an aspect ratio of height/width of 1:1 and stroke width 1/5 of height/width. Each optotype is surrounded by crowding bars with same stroke width of 1/5 height/width and an angular separation of ½ height/width from the optotype. EVA uses an automated non-modifiable ATS protocol (fully described in EVA Procedure section). The system is calibrated for a 3 meter distance. The luminance was measured as 40.4 ft-cd (138 cd/m^2) and Michelson contrast was 0.869 (Weber contrast 0.930) as measured by a Spectra Pritchard Photometer at a distance of 1.5 meters with aperture size of 6 minarc.

Figure 5: EVA



III. Experimental procedure

A brief history and informed consent from the subjects were obtained on the first visit. The habitually-corrected right eye of each subject was tested; the left eye was occluded with an adhesive patch during testing.

There were three visual acuity devices presented to each subject: a standard printed HOTV chart in logMAR format (see Section II.1.), a comparison device, Smart System II PC (SS II) (see Section II.2) and second comparison device, the Electronic Visual Acuity Tester (EVA) (see Section II.3). To test the validity of the comparison devices, the threshold for resolution of optotypes was calculated and compared to the threshold obtained with the logMAR HOTV chart. All charts were tested in the same exam room with an ambient illumination of 340 lux (32ft-cd) and were presented at a test distance of 4 meters (except for EVA, where the test distance was 3 m (see "EVA Procedure" section).

To investigate test-retest reliability for each comparison device, visual acuity measurements were taken in two visits separated by approximately 1-2 weeks. In order to prevent memorization and fatigue, the order of presented visual acuity charts was randomized. The order of the test devices was reversed on the second visit. Both visits were conducted in the same exam room on the third floor of 418 Beacon Street with standard overhead fluorescent lighting and previously mentioned ambient illumination.

On each visit, the 30 subjects were tested with all three devices under two optical conditions: with habitual-correction and with myopic blur. Myopic blur was induced by placing +1.25 DS lens over the habitually-corrected right eye. The order of habitually corrected and blur-induced conditions was randomized and there was a two minute break

between each experiment to ensure that the patient has time to adjust to/from the induced blur. The order of these conditions was reversed on the second visit.

1. HOTV chart test procedure:

The logMAR HOTV chart has all visual acuity lines presented on the screen at once. The chart uses the denominator progression in the Snellen equivalent fraction 20/x of 100, 80, 63, 50, 40, 32, 25, 20, 16, 12.5 and 10 (corresponding to a 0.1 logMAR progression). Each subject completed visual acuity measurement by reading 5 optotypes on each line starting with a +0.70 logMAR visual acuity line and until a termination criterion occurred. The termination rule used in the study was designed to conform to clinical practice by discontinuing testing on the line at which four optotypes are incorrectly identified. Once termination criterion is reached, threshold is established as described in "Threshold Calculation" section.

***2. SSII test procedure*:**

The SS II device presents isolated crowded optotypes one at a time in a random order with a total of five optotypes presented for each acuity level. The system uses the denominator progression in Snellen equivalent fraction 20/x of 100, 80, 60, 50, 40, 30, 25, 20, 15 and 10. Each subject completed visual acuity measurement by reading 5 optotypes on each level starting with the 20/100 line and until a termination criterion occurred. The same termination rule was used as for the logMAR HOTV test.

***3. EVA test procedure*:**

EVA device uses the automated Amblyopia Treatment Study (ATS protocol) described in Table 1. The protocol is calibrated for a 3-meter test distance (versus 4 meter test distance used with logMAR HOTV chart and SS II system), presenting isolated crowded optotypes one at a time in a randomized order. EVA uses the denominator progression in the Snellen equivalent fraction 20/x of 100, 80, 63, 50, 40, 32, 25, 20 and 16 (which is a 0.1 logMAR progression). Each subject completed visual acuity measurement using the automated ATS protocol starting from 20/100 level, using an adaptive staircase procedure which varies stimulus size and number of presentations depending upon the subject's responses.

Table 1: EVA Test Procedure: Amblyopia Treatment Study Visual Acuity Protocol

Single letters with surround bars are presented in four phases: screening, phase I (first threshold determination), reinforcement, and phase II (second threshold determination). • In phase I and phase II, up to four single letters are sequentially presented at each logMAR level* that is tested. • A level is considered to be 'passed' if 3 of 3 or 3 of 4 letters are correct and 'failed' if 2 letters at a level are missed. • Testing of a level stops as soon as criteria are met for either 'pass' or 'fail'.
Screening Phase *Starting from either 20/100 or 20/400, single letters, in sequential descending logMAR sizes, are shown until one is missed.* 1. Tester selects 20/100 or 20/400 size letter to present as starting point (depending on the expectation of visual acuity level based on previous testing or a pretest).

2. If response is correct, letter at next smallest logMAR level is presented and testing continues sequentially with one letter per logMAR level through 20/20 until there is an incorrect response.

   If starting point was 20/100 and response is incorrect at either 20/100 or 20/80, screening is restarted at 20/400.

4. If screening starts at 20/400 and response is incorrect, testing continues sequentially with one letter per logMAR level through 20/800 until there is a correct response.

   - If 20/800 is missed, 20/800 becomes the starting level for phase I.

Phase I

Starting 2 logMAR levels above the missed level in Screening, the smallest logMAR level at which 3 of 3 or 3 of 4 letters are correctly identified is determined.

1. Up to 4 single letters are sequentially presented 2 logMAR levels above the level missed in screening
   - Exception: if 20/20 was correct in Screening, phase I starts at 20/30.
   - Exception: if 20/800 or 20/640 was missed in Screening, phase I starts at 20/800
2. If first tested level is failed, testing continues at sequentially larger logMAR levels until a level is passed.
   - If 20/800 is failed, phase I ends, the Reinforcement phase is omitted, and 20/800 is retested in phase II
3. If first tested level is passed, testing continues at sequentially smaller logMAR levels until a level is failed.

Reinforcement Phase

In order to get the child whose attention is drifting back on track, 3 letters larger than the phase I threshold are sequentially presented.

1. Starting 3 levels larger than the level missed in phase I, 3 successively smaller single letters are presented.
   - Exception: if the level failed in phase I is 20/500 or 20/640, 3 20/800 letters are shown for reinforcement

Note: the reinforcement phase responses do not contribute to the visual acuity score and even if the responses are incorrect, the test proceeds to phase II.

Phase II

The last level missed in Phase I is retested and if 'passed', testing continues until a level is 'failed'

1. Up to 4 single letters are sequentially presented at the last level missed in phase I
2. If the level is failed, testing stops.
3. If the level is passed, testing continues at sequentially smaller logMAR levels until a level is failed.

Final Visual Acuity Score
The visual acuity score is the smallest logMAR level passed in Phase I or Phase II.
*logMAR levels (Snellen equivalents): 20/800, 20/640, 20/500, 20/400, 20/320, 20/250, 20/200, 20/160, 20/125, 20/100, 20/80, 20/63, 20/50, 20/40, 20/32, 20/25, 20/20, 20/16

There were two differences between the presentation of the logMAR HOTV chart and SS II stimuli:

1) The logMAR HOTV chart presents a full chart with 5-per line optotypes in lines in logMAR progression. Only one version of the logMAR HOTV chart was used for testing. Whereas, SS II presents a single optotype one at a time in a random sequence preventing immediate optotype repetition.

2) The logMAR HOTV chart identifies each line by the denominator of the Snellen equivalent fraction (20/x) as 100, 80, 63, 50, 40, 32, 25, 20, 16, 12.5 and 10. SS II, on the other hand, uses the denominator progression of 100, 80, 60, 50, 40, 30, 25, 20, 15 and 10 with a missing 12.5 line and Snellen progression of optotypes sizes, which is a potential source of small error for a letter-by-letter calculation of threshold acuity.

There were four differences between the logMAR HOTV chart and EVA system:

1) logMAR HOTV chart presents a full chart with all optotypes in lines, 5 per line (exept 20/100), in logMAR progression. Only one version of the HOTV chart was used for testing. Whereas, EVA presents single optotypes in a random sequence without preventing immediate optotype repetition.

2) The logMAR HOTV chart includes 20/12.5 and 20/10 lines while the smallest acuity line in EVA is 20/16. This could cause underestimation of scored visual acuity for subjects whose visual acuity is better than 20/16.

3) Up to four letters are presented at each LogMAR level during the testing phases with the EVA system. By contrast, with the logMAR HOTV chart all 5 letters on each acuity line were tested.

4) Following the EVA ATS test protocol, test distance is 3 meters compared to the standard 4-meter distance for the standard HOTV chart.

IV. Threshold Calculation

To compare acuity measurements between the three tests, ideally the same cut-off criterion should be used. Two main threshold scoring methods have been reported: line-by-line (Pass/Fail) and letter-by-letter (Interpolation). Both methods are in wide use in clinical adult and pediatric research with little agreement about the best scoring method to use with logMAR charts. Studies have shown that letter-by-letter scoring gives a finer gradation of measurement and lower variance than does a Pass/Fail method (Carkeet, 2001; Raasch and Bailey, 1998). However, other studies have reported small differences between the two threshold scoring methods (Carkeet, 2001; Chen et al, 2006). There is also little agreement in the literature as to the best termination rule to use. In this study, the threshold criterion was chosen as 60%-correct, that is, the smallest line where at least three out of five optotypes

were identified correctly. Studies have shown reliable results when the test is stopped after three or more mistakes (Brown and Lovie-Kitchin, 1993; Carkeet, 2001).

Letter-by-letter scoring presumes that, when comparing different tests, the sequence of optotype sizes and the total number of optotypes shown are equivalent across tests. This turned out *not* to be true for this study. SS II does not test 20/12.5 visual acuity line which results in a different total number of letters presented to the subject when compared to the logMAR HOTV chart. For the same reason, a Pass/Fail scoring method has limited use for SS II since it can potentially create a clinically significant error in threshold calculation as the gap between 20/16 and 20/10 line is 0.2 LogMAR units. In addition, SS II stimuli were not spaced in equal logMAR intervals as in HOTV chart. So the comparison of SS II to the HOTV 0.1 logMAR interval acuity scale cannot be exact, unavoidably. However, the difference is likely to be small. In this study the assumption was made that logMAR HOTV chart and SS II device have approximately equal logMAR differences between acuity levels. Thus, a threshold can be calculated if the interval between the line above and below the threshold is known. Ideally, acuity would be scored on each test as the level at which subject judged 3 of 5 optotypes correctly, based on the 60%-correct threshold criterion. To come up with an equivalent threshold criterion for the SSII subjects whose threshold was better than 20/16, an “Interpolation” procedure was utilized to determine 60%-correct (threshold). This method of threshold calculation is labeled Interpolation of Pass/Fail method and is based on the interpolated logMAR value corresponding to 60% correct performance.

Description of Interpolation of Pass/Fail Method:

1) The smallest optotype line (or acuity level) with three correctly named optotypes out of five presented (3/5 correct) was scored at exactly that acuity level (logMAR value);

2) If 3/5 correct was not reached on the line below 4/5 or 5/5 correct (ie. 2/5, 1/5 or 0/5 correct), the "true" threshold was calculated interpolating between the line above threshold criterion (4/5 or 5/5 correct) and the line below threshold (2/5, 1/5 or 0/5 correct) as follows:

$$VA_I = \log \frac{10^{\log(B)+\left((\log(A)-\log(B))\bullet\frac{p(t)-p(b)}{p(a)-p(b)}\right)}}{20}$$

VA_I = visual acuity as estimated by the interpolation method (in logMAR units)

A = Snellen equivalent of visual acuity line above visual acuity threshold

B = Snellen equivalent of visual acuity line below visual acuity threshold

p(t) = percent recognized letters that defines the visual acuity threshold

p(b) = percent recognized letters below the visual acuity threshold

p(a) = percent recognized letters above the visual acuity threshold

Please note: when visual acuity is identical and above threshold in two lines (i.e. p(a) = p(b) = 0.6) VA_I is calculated as:

$$VA_I = \frac{\log(A) + \log(B)}{2}$$

3) Cases where there was a reversal of correct scores (larger optotype lines with 3/5 correct, and smaller optotypes lines with 4/5 or 5/5 correct), threshold was interpolated using the finest optotype acuity with 4/5 or 5/5 correct respectively, ignoring the 3/5 correct line above the 4/5 or 5/5 correct lines;

4) In cases of multiple 3/5 correct lines, the logarithmic average was taken between the lines. Interpolation of Pass/Fail threshold scoring method was used for both tests, logMAR HOTV chart and SS II, in order to be able to directly compare the visual acuity measurements between the charts.

EVA follows an automated scoring procedure that differs from the standard ETDRS protocol (Table 1). EVA uses a Pass/Fail method with the threshold criterion of 75%-correct that is the smallest line where at least three out of four optotypes were identified correctly. In contrast, the logMAR HOTV presents 5 optotypes at each visual acuity level, and a threshold criterion was chosen as 3-correct (60%). This criterion is frequently employed in clinical research and practice, thus ensuring generalizability of our findings. Therefore, to ensure a comparable method of visual acuity calculation, Pass/Fail threshold calculation method was chosen to compare validity of EVA results to logMAR HOTV chart.

Threshold Definitions

1. logMAR HOTV chart threshold calculation

Interpolation of Pass/Fail threshold scoring method was used to compare visual acuity measurements with SS II. Pass/Fail 60%-correct method was used to compare visual acuity measurements with EVA.

2. ***SS II threshold calculation:*** Interpolation of Pass/Fail threshold scoring method was used.

3. ***EVA threshold calculation:*** an automated ATS protocol was used that utilizes a Pass/Fail method with the threshold criterion of 75%-correct.

V. Data Analysis

The test results were recorded by hand on a miniature recording sheet emulating each acuity chart. The data were then entered into a database as a number correct for each visual acuity line. Acuities were converted to logMAR scale. Bland-Altman method (Bland and Altman, 1986) and paired two-sample t-test were utilized in the statistical analysis. A Bland-Altman analysis was used for the results from each test device to establish 95% limits of agreement for test-retest repeatability. Agreement between tested computerized methods and HOTV reference in estimating visual acuity was further quantitatively explored using Bland-Altman plots. Comparison of 95% limits of agreement (LOA) between the devices was also performed for completeness (Altman and Bland, 1983; Bland and Altman, 1986; Neuhäuser and Manly, 2004). Overlapping values for individual observations were plotted at identical coordinates in these plots, but all plots presented in the present study include all collected data points (n = 30). Validity analyses (i.e. comparison of VA estimated by SS II or EVA vs. logMAR HOTV) were performed using only data from week 1. All thresholds for statistical significance were set at $p < 0.05$ (two-tailed). All analyses were performed using R v 2.10.0 (The R Statistical Computing Corporation).

Analysis Plan

1. ***Comparison of Interpolation of Pass/Fail method with standard Pass/Fail method for threshold calculations using logMAR HOTV data.*** Bland-Altman method and paired two-sample t-test were utilized in the statistical analysis. One analysis was done comparing the acuity results from 30 subjects using the logMAR HOTV chart

for Week 1 visit using habitual correction conditions and a separate analysis was performed for myopic blur-induced conditions.

2. ***SS II: validity and repeatability under habitually-corrected condition and under myopic-blur induced condition.*** A Bland-Altman analysis and paired two-sample t-tests were utilized in the statistical analysis. These analyses were done comparing visual acuity threshold of 30 subjects for Week 1 visit. Separate analyses was performed for myopic blur-induced conditions using the results of the same 30 subjects' Week 1 visit.

3. ***EVA: validity and repeatability under habitually-corrected condition and under myopic-blur induced condition.*** The same analyses were conducted as with SSII data.

RESULTS

Table 2. Summary of Analyses

Devices Tested	Testing Conditions	Comparison	Testing Purpose
HOTV	Habitually-corrected	Interpolation method vs. Pass/Fail Method	Assess comparability of different threshold calculation methods
HOTV	Blur-induced	Interpolation method vs. Pass/Fail Method	
HOTV, SS II	Habitually-corrected	HOTV: VA week 1 vs. VA week2	Compare test-retest repeatability of HOTV and SS II
		SS II: VA week 1 vs. VA week2	
HOTV, SS II	Blur-induced	HOTV: VA week 1 vs. VA week2	
		SS II: VA week 1 vs. VA week2	
HOTV, EVA	Habitually-corrected	HOTV: VA week 1 vs. VA week2	Compare test-retest repeatability of HOTV and EVA
		EVA: VA week 1 vs. VA week2	
HOTV, EVA	Blur-induced	HOTV: VA week 1 vs. VA week2	
		EVA: VA week 1 vs. VA week2	
HOTV, SS II	Habitually-corrected	SS II VA week 1 vs. HOTV VA week 1	Assess validity of SS II compared to HOTV chart
HOTV, SS II	Blur-induced	SS II VA week 1 vs. HOTV VA week 1	
HOTV, EVA	Habitually-corrected	EVA VA week 1 vs. HOTV VA week 1	Assess validity of EVA compared to HOTV chart

HOTV, EVA	Blur-induced	EVA VA week 1 vs. HOTV VA week 1	

I. Comparison of Interpolation of Pass/Fail method with standard Pass/Fail method of threshold calculation.

To examine the validity of the Interpolation of Pass/Fail method, acuity data were compared between the standard Pass/Fail threshold criteria (ignoring reversals) and the Interpolation of Pass/Fail method using results of logMAR HOTV test week 1. For both blurred and myopic blur-induced conditions, statistical analyses showed a significant difference when comparing mean values for Interpolation of Pass/Fail and Pass/Fail method ($p < 0.05$). The absolute average difference between methods in visual acuity estimation was 0.03 logMar (equivalent to 1.5 optotypes) for both comparisons (Figs. 6 and 7). The 95% limits of agreement between methods ranged from -0.0825 to +0.0211 logMAR units for habitually-corrected conditions and from -0.0881 to +0.0251 for myopic-blur induced conditions. Overall, the Interpolation of Pass/Fail method resulted in consistently lower (better) visual acuity than the Pass/Fail criterion ($p < 0.05$), as expected based on the differences in threshold calculation, but this systematic difference was found to be independent of subjects' gender, visual acuity, testing session and condition (habitually-corrected vs. myopic-blur induced) (all $p > 0.20$).

Figure 6: Differences in visual acuity estimation for the HOTV chart comparing Interpolation of Pass/Fail (Interpol LogMar) and Pass/Fail (P/F) threshold method under habitually-corrected conditions

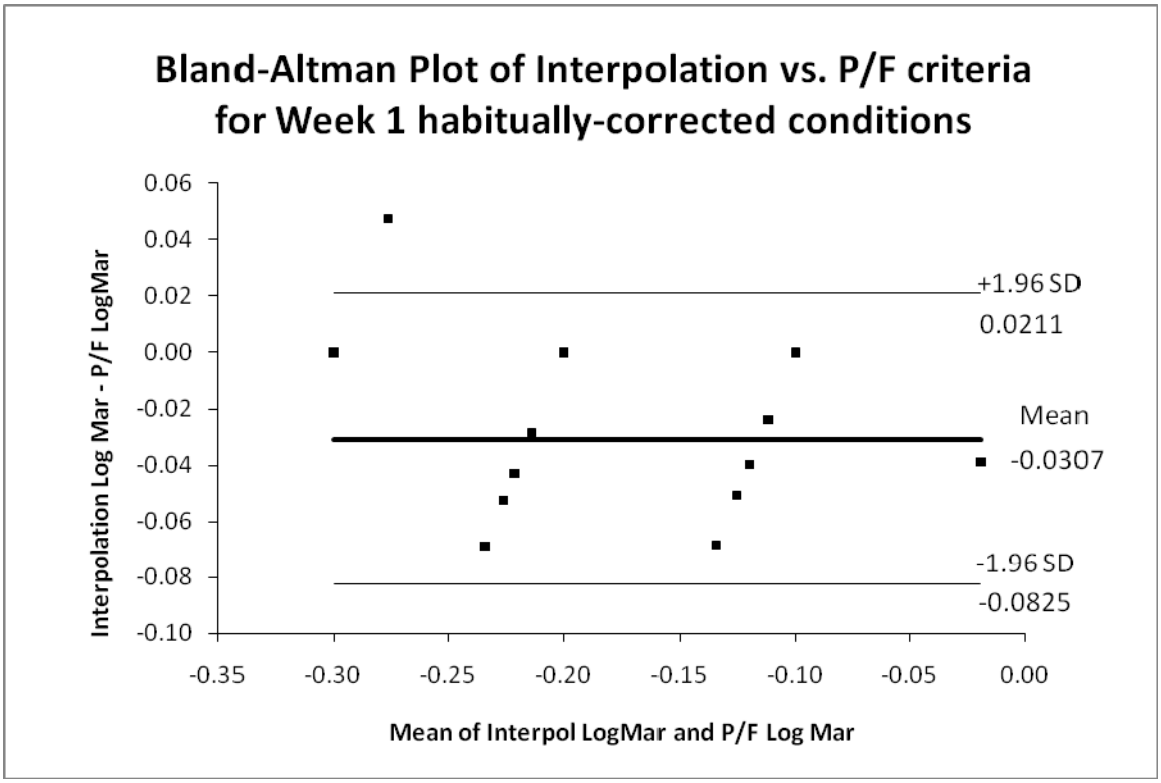
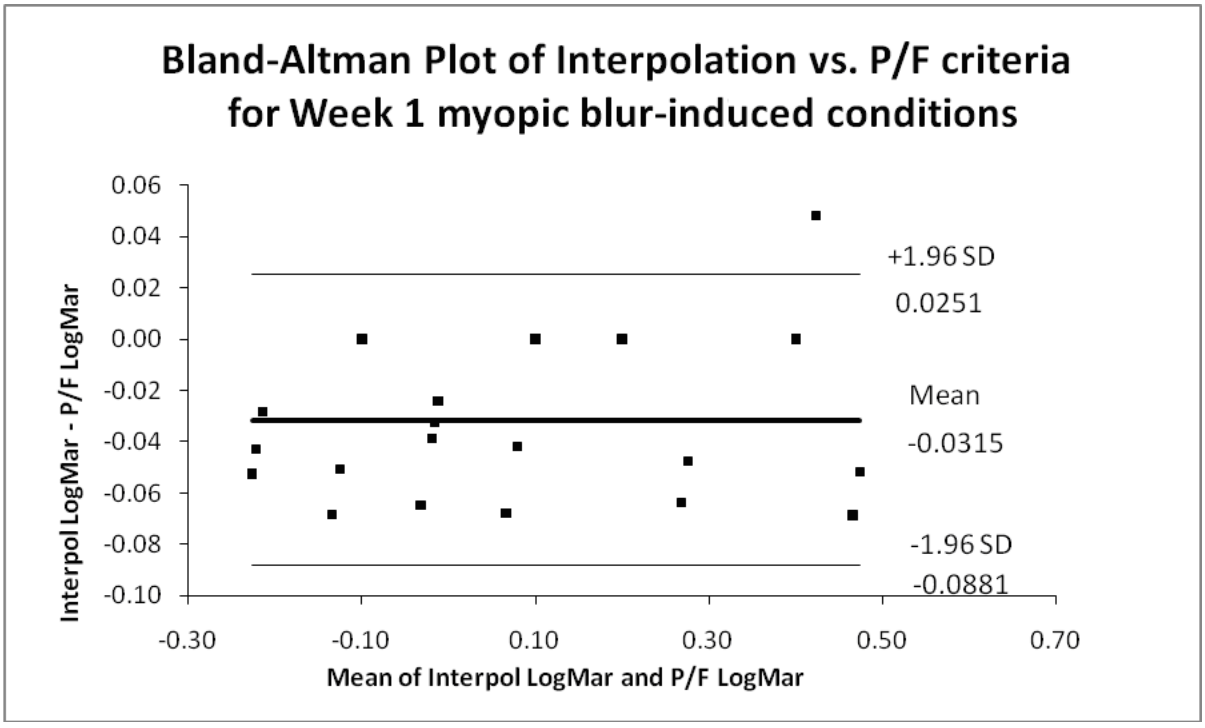


Figure 7: Differences in visual acuity estimation for the HOTV chart comparing Interpolation of Pass/Fail (Interpol LogMar) and Pass/Fail (P/F) threshold method under myopic blur-induced conditions



II. SS II: repeatability under habitually-corrected and myopic-blur induced conditions

Under habitually corrected conditions, the mean visual acuity tested by SS II was -0.186 ± 0.068 logMAR on week 1 -0.165 ± 0.066 logMAR on week 2 (Table 3). Tested with the Goodlite HOTV chart, mean visual acuity was -0.204 ± 0.076 logMAR on week 1 and -0.210 ± 0.078 one week 2. There was no statistically significant difference between the results of week 1 and week 2 for both SS II and logMAR HOTV tests (Table 3). For SSII the 95% limits of agreement for visual acuity ranged from +0.112 to -0.152 logMAR units. In comparison, results from logMAR HOTV chart showed 95% limits of agreement from +0.067 to -0.062 logMAR (Table 3). No statistically significant difference was found when comparing mean difference and limits of agreement in repeatability of visual acuity estimation for SSII vs. logMAR HOTV ($p > 0.05$).

Under **blur-induced conditions**, mean visual acuity tested by SS II was +0.044 ± 0.189 logMAR on week 1 and +0.040 ± 0.195 logMAR on week 2 (Table 3). Tested with logMAR HOTV chart, mean visual acuity was +0.048 ± 0.220 logMAR on week 1 and +0.017 ± 0.210 logMAR on week 2. No statistically significant difference between the results of week 1 and week 2 was found for both SS II and logMAR HOTV tests (Table 3). For SSII the 95% limits of agreement for visual acuity ranged from +0.330 to -0.321 logMAR. In comparison, results from logMAR HOTV chart demonstrated 95% limits of agreement ranging from +0.184 to -0.152 logMAR units (Table 3). No statistically significant difference

was found when comparing the mean and 95% limits of agreement for agreement in estimation for SSII vs. logMAR HOTV ($p > 0.05$).

Table 3: Repeatability of visual acuities for Goodlite HOTV and SSII

Testing procedure	**Condition**	**VA* Week 1**	**VA* Week 2**	**p§**	**Bland-Altman Analysis†**	
					Mean difference	***95% LOA***
HOTV	Habitual	-0.204 (0.076)	-0.210 (0.078)	0.63	0.003	- 0.062 /+0.067
SS II	Habitual	-0.186 (0.068)	-0.165 (0.066)	0.11	-0.020	-0.152 / +0.112
HOTV	Blur	0.048 (0.220)	0.017 (0.210)	0.32	0.016	-0.152 / +0.184
SS II	Blur	0.044 (0.189)	0.040 (0.195)	0.88	0.004	-0.321 / +0.330

* Reported as mean (SD) visual acuity in logMar units.

§ p-value based on two-tail paired two-sample t-test analysis

† values presented refer to mean and 95% CI difference in VA acuity estimation (in logMar units) comparing testing sessions

95% LOA = 95% Limits of agreement, p = p-value, VA = visual acuity

III. EVA: repeatability under habitually-corrected and myopic-blur induced conditions

Under habitually corrected conditions, mean visual acuity as tested by EVA was -0.087 ± 0.035 logMAR on week 1 and -0.097 ± 0.018 logMAR units on week 2 (Table 4). In comparison, if tested with logMAR HOTV chart, Pass/Fail mean visual acuity was found to be -0.173 ± 0.082 logMAR units on week 1 and -0.174 ± 0.094 logMAR units on week 2. Statistical analysis found no significant difference between results of week 1 and week 2 for both EVA and logMAR HOTV (Table 4). For EVA the 95% limits of agreement for visual acuity ranged from +0.070 to -0.050 logMAR units. In comparison, testing with the HOTV chart demonstrated 95% limits of agreement ranging from +0.162 to -0.160 logMAR units (Table 3). Differences in agreement in visual acuity estimation comparing different testing sessions were significantly larger for logMAR HOTV compared to EVA ($p < 0.05$).

Under **blur-induced conditions**, mean visual acuity for EVA was +0.083 ± 0.158 logMAR on week 1 and +0.087 ± 0.185 logMAR on week 2 (Table 4). In comparison, logMAR HOTV chart's mean visual acuity was found to be +0.080 ± 0.212 during week 1 and +0.049 ± 0.207 during week 2. No significant difference was found when comparing results of week 1 and week 2 for both EVA and logMAR HOTV (Table 4). The 95% limits of agreement for visual acuity ranged from +0.163 to -0.170 logMAR units for EVA. In comparison, analysis of logMAR HOTV chart results returned 95% limits of agreement ranging from +0.384 to -0.316 logMAR units (Table 4): limits of agreement were significantly larger for logMAR HOTV compared to EVA ($p < 0.05$).

Table 4: Repeatability of visual acuities for logMAR HOTV and EVA

Testing procedure	Condition	VA* Week 1	VA* Week 2	p§	Bland-Altman Analysis†	
					Mean difference	*95% LOA*
HOTV	Habitual	-0.173 (0.082)	-0.174 (0.094)	0.96	0.001	-0.160 / +0.162
EVA	Habitual	-0.087 (0.035)	-0.097 (0.018)	0.083	0.010	-0.050 / +0.070
HOTV	Blur	0.080 (0.212)	0.049 (0.207)	0.30	0.034	-0.316 / 0.384
EVA	Blur	0.083 (0.158)	0.087 (0.185)	0.83	-0.003	-0.170 / +0.163

* Reported as mean (SD) visual acuity in logMar units.

§ p-value based on two-tail paired two-sample t-test analysis

† values presented refer to mean and 95% LOA in VA acuity estimation (in logMar units) comparing testing sessions

95% LOA = 95% Limits of agreement, p = p-value, VA = visual acuity

IV. SS II: validity under habitually-corrected and myopic-blur induced conditions

Under habitually corrected conditions, there was no statistically significant difference in mean visual acuity between SS II and logMAR HOTV charts ($p = 0.12$). Bland-Altman plot analysis was used to quantitatively present the difference of agreement between the instruments (Table 5 and Figure 8). The mean visual acuity difference between charts was -0.018 ± 0.063, with SSII showing a non-significant trend towards underestimation of visual acuity. 95% limits of agreement in visual acuity estimation between the two instruments ranged from +0.104 to -0.141 logMAR. Under myopic blur-induced conditions, there was no significant difference between mean visual acuities of SS II and HOTV charts ($p = 0.90$). Agreement in visual acuity estimation between the instruments is presented in the Bland-Altman plot analysis (Table 5 and Figure 9). 95% Limits of agreement between the two instruments ranged from +0.353 to -0.345 logMAR.

Table 5: Validity of SS II VA measurements as compared to HOTV

Testing Condition	Mean VA* HOTV	Mean VA* SSII	p§	Bland-Altman Analysis†	
				Mean difference	***95% LOA***
Habitual	-0.204 (0.076)	-0.186 (0.068)	0.12	-0.018	-0.141 / + 0.104
Blur	+0.048 (0.220)	+0.044 (0.189)	0.90	0.004	-0.345 / +0.353

* Reported as mean (SD) visual acuity in LogMar using the Interpolation method

§ p-value based on two-tail paired two-sample t-test analysis

† values presented refer to mean and 95% LOA in VA acuity estimation (in LogMar units) comparing different instruments

95% LOA = 95% Limits of agreement, p = p-value, VA = visual acuity

Figure 8: Bland-Altman plot of visual acuities for logMAR HOTV vs. SSII under habitual corrected conditions (n = 30)

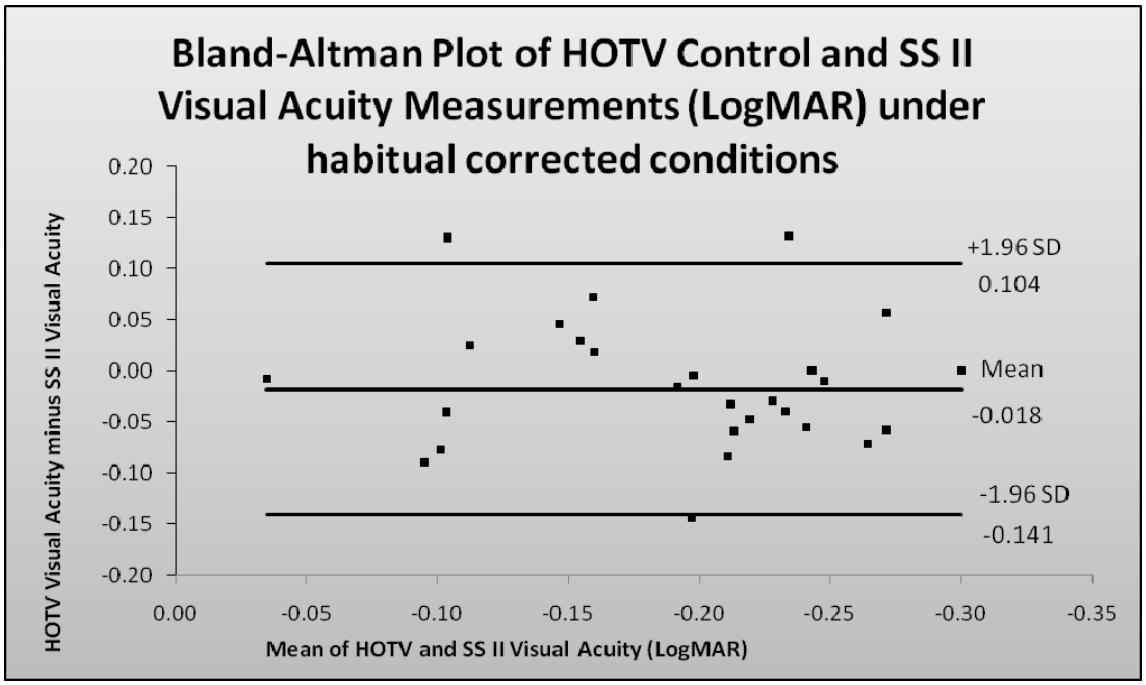
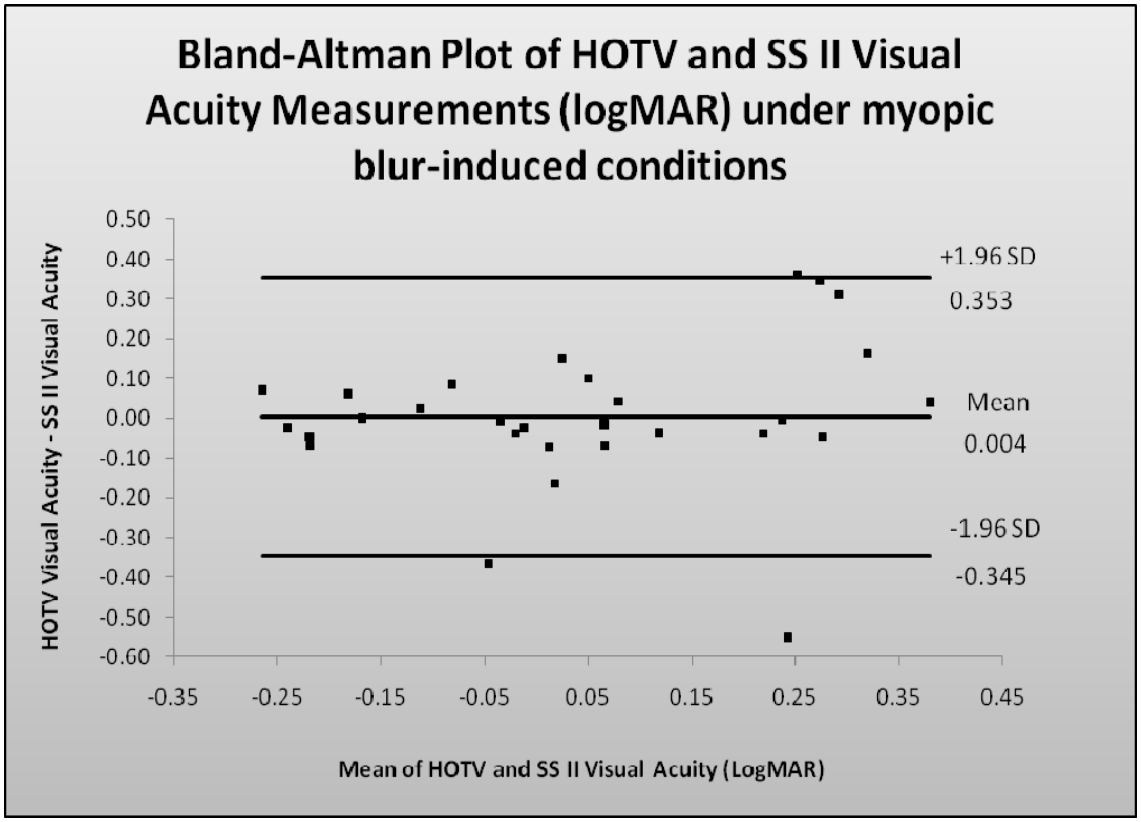


Figure 9: Bland-Altman plot of visual acuities for logMAR HOTV vs. SSII under blur-induced conditions (n = 30)



V. EVA: validity under habitually-corrected and myopic-blur induced conditions

Under habitually-corrected conditions EVA underestimated visual acuity measurements (as compared to logMAR HOTV chart) by a mean value of -0.087 logMAR ($p < 0.05$) (Table 6). The 95% limits of agreement between the two instruments ranged from +0.056 to -0.230 logMAR, with the results being skewed towards underestimation of LogMAR visual acuity (Figure 10, all n = 30 observations presented in the graph; identical results plotted as overlapping values as described in the Methods section). Under myopic-blur induced conditions, there was no significant difference between mean visual acuities comparing results for EVA and logMAR HOTV charts (Table 6). The limits of agreement between the two instruments ranged from +0.246 to -0.253 logMAR (Figure 11, all n = 30 observations presented in the graph; identical results plotted as overlapping values as described in the Methods section).

Table 6: Validity of EVA visual acuity compared to logMAR HOTV visual acuity

Testing Condition	Mean VA* HOTV	Mean VA* SSII	p§	Bland-Altman Analysis†	
				Mean difference	*95% LOA*
Habitual	-0.173 (0.082)	-0.087 (0.035)	< 0.05	-0.087	-0.230 / +0.056
Blur	-0.080 (0.212)	-0.083 (0.158)	0.89	+0.003	-0.253 / +0.246

* Reported as mean (SD) visual acuity in LogMar units using the Pass/Fail method

§ p-value based on two-tail paired two-sample t-test analysis

† values presented refer to mean and 95% CI difference in VA acuity estimation (in LogMar units) comparing different instruments

95% LOA = 95% Limits of agreement, p = p-value, VA = visual acuity

Figure 10: Bland-Altman plot of visual acuities for HOTV vs. EVA under habitual corrected conditions (n = 30)

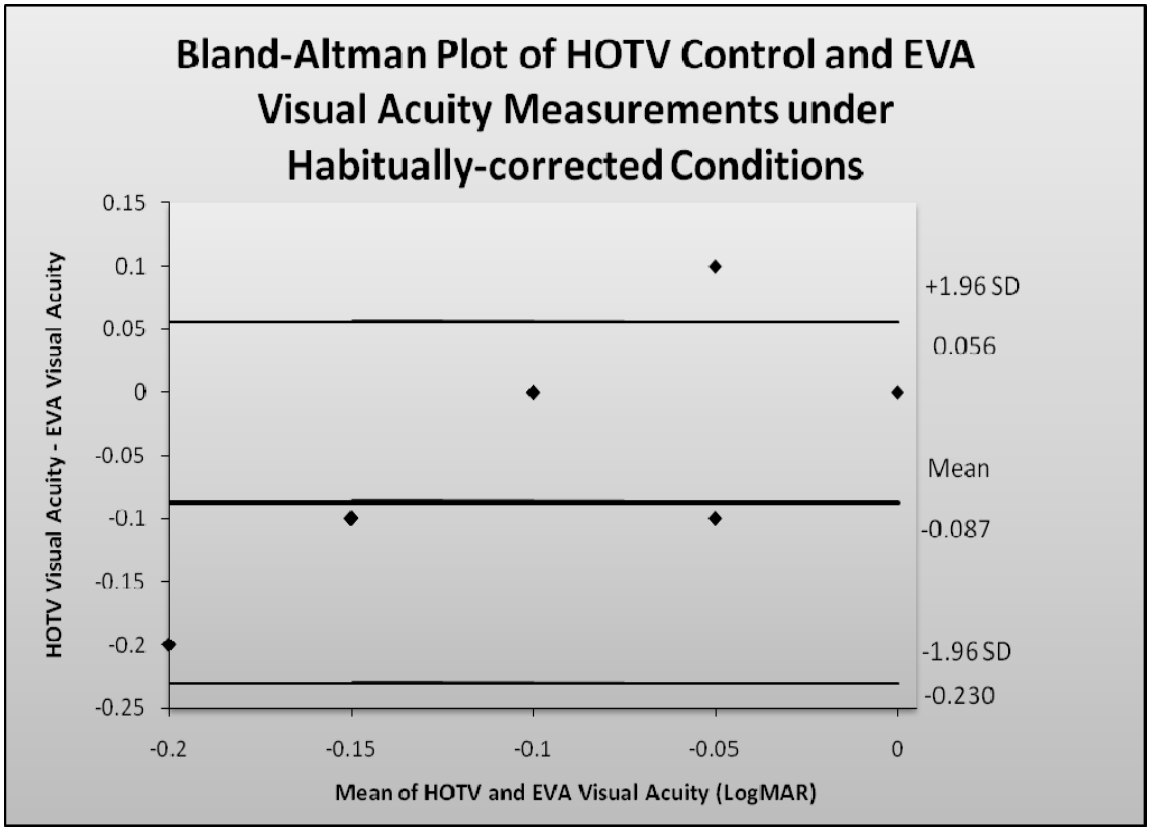
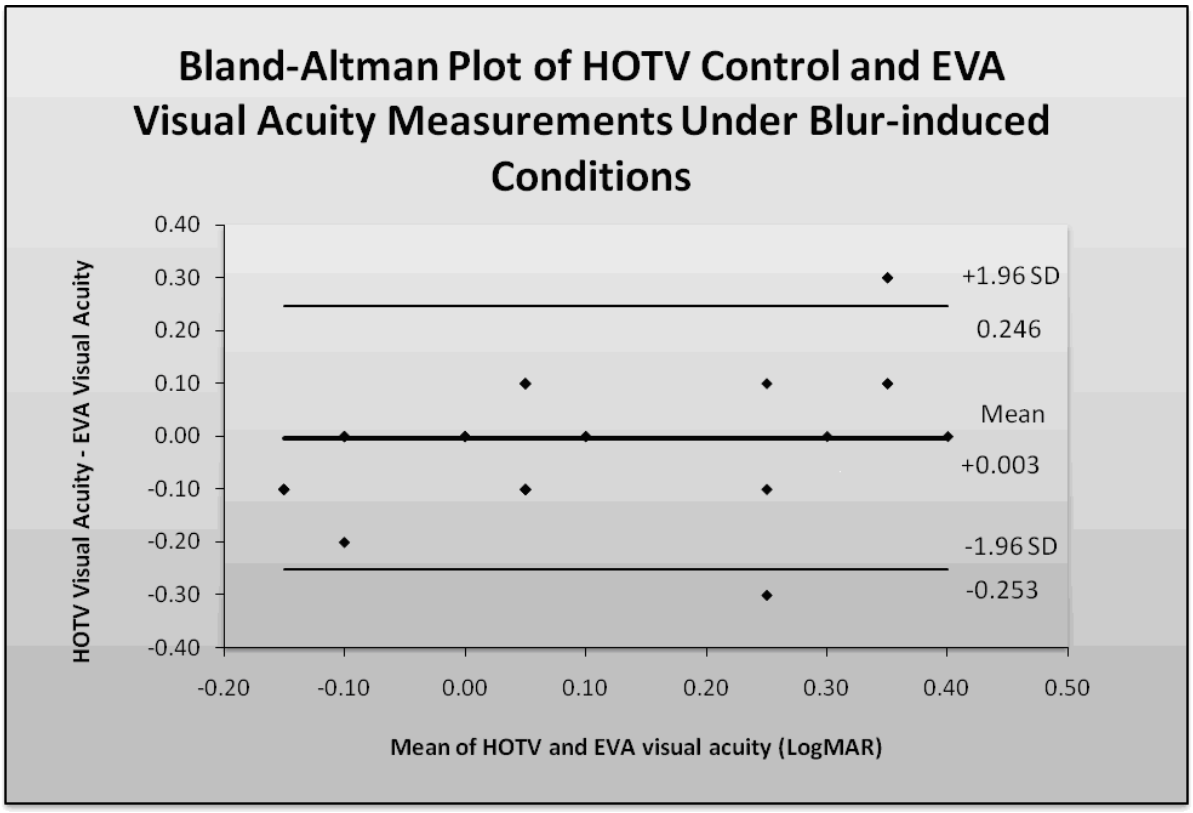


Figure 11: Bland-Altman plot of visual acuities for HOTV vs. EVA under blur-induced conditions (n = 30)



DISCUSSION AND CONCLUSIONS

Summary of Findings

The results from the present study indicate that both computerized systems studied (SS II and EVA) showed no statistically significant differences in visual acuities between data obtained at different short time points. That is, there were no differences in test-retest reliability between SSII or EVA. The 95% limits of agreement for visual acuity across different testing sessions were comparable for HOTV and SS II. However, 95% limits of agreement were significantly larger for logMAR HOTV compared to EVA ($p < 0.05$) for both optical conditions, suggesting that test-retest reliability (or repeatability) is poorer for logMAR HOTV vs. EVA. SSII and HOTV 95% limits of agreement were similar. While SS II visual acuity showed concurrent validity in comparison with standard logMAR HOTV visual acuity in all testing conditions, EVA visual acuity underestimated logMAR HOTV visual acuity under habitually-corrected conditions. All results from the present study and their interpretation are summarized in Table 7.

Table 7. Summary and Interpretation of Findings

Devices Tested	Testing Conditions	Comparison	Testing Purpose	Interpretation of Results
HOTV	Habitually-corrected	Interpolation method vs. Pass/Fail Method	Assess comparability of different threshold calculation methods	Difference in VA estimated using different methods, but limited clinical relevance (< 1 line on logMAR chart)
HOTV	Blur-induced	Interpolation method vs. Pass/Fail Method		
HOTV, SS II	Habitually-corrected	HOTV: VA week 1 vs. VA week2	Compare repeatability of HOTV and SS II	No difference in repeatability (mean VA and 95% LOA) comparing HOTV and SS II
		SS II: VA week 1 vs. VA week2		
HOTV, SS II	Blur-induced	HOTV: VA week 1 vs. VA week2		
		SS II: VA week 1 vs. VA week2		
HOTV, EVA	Habitually-corrected	HOTV: VA week 1 vs. VA week2	Compare repeatability of HOTV and EVA	Difference in repeatability comparing HOTV and EVA: better repeatability for EVA under both testing conditions ($p<0.05$)
		EVA: VA week 1 vs. VA week2		
HOTV, EVA	Blur-induced	HOTV: VA week 1 vs. VA week2		
		EVA: VA week 1 vs. VA week2		
HOTV, SS II	Habitually-corrected	SS II VA week 1 vs. HOTV VA week 1	Assess validity of SS II compared to HOTV chart	No difference in VA estimation between HOTV and SS II
HOTV, SS II	Blur-induced	SS II VA week 1 vs. HOTV VA week 1		
HOTV, EVA	Habitually-corrected	EVA VA week 1 vs. HOTV VA week 1	Assess validity of EVA compared to HOTV chart	Difference in VA estimation between HOTV and EVA under habitually-corrected conditions only, ($p< 0.05$), but limited clinical relevance (< 1 line on logMAR chart). No difference under blur-induced conditions
HOTV, EVA	Blur-induced	EVA VA week 1 vs. HOTV VA week 1		

Threshold calculation

Comparison of different VA testing systems in the present study required application of an interpolation method for threshold calculation, as opposed to a pass/fail scoring criterion alone. A comparative analysis of these methods identified a significant difference in mean visual acuity: as expected, the interpolation method resulted in consistently lower (better) visual acuity estimates when analyzing the HOTV data. However, both the absolute mean difference in visual acuity and the limits of agreement when comparing methods were lower than the smallest reliably identifiable changes in visual acuity in clinical practice (~0.10 logMAR units, or a single line), based on a single study (Arditi and Cagenello, 1998). Furthermore, differences in visual acuities between methods were not dependent on optical testing condition, supporting the hypothesis of our choice of threshold definition not exerting undue influence on overall findings.

SS II: repeatability

Habitually-corrected conditions: No statistically significant difference was found between mean visual acuity on weeks 1 and 2 ($p>0.05$) under habitually-corrected conditions when tested with SS II. The 95% limits of agreement for SS II when comparing week 1 to week 2 ranged from +0.112 to -0.152 logMAR. In comparison, the logMAR HOTV chart showed somewhat smaller limits of agreement when comparing week 1 to week 2. Comparison of limits of agreement between logMAR HOTV vs. SS II was not statistically significant, although it did show a trend towards difference in test-retest reliability ($p = 0.12$). However, the observed disparities in repeatability (HOTV week 1 vs. week 2: +0.003

logMAR units; SS II week 1 vs. week 2: -0.020 logMAR units) are very small (~ 1 optotype) and unlikely to be of clinical relevance.

Limits of agreement for test-retest reliability in our study are consistent with previously reported repeatability results for crowded logMAR visual acuity in pediatric and adult populations (Table 8). In the present study and that of Manny et al (Manny et al, 2003), normal subjects were tested with logMAR visual acuity chart with comparable repeatability results. Similar limits of agreement for different testing sessions were also reported by Hazel and Elliott (Hazel and Elliott, 2002) (~ ±0.14 logMAR). One study compared a computerized visual acuity test device similar to SSII (by M&S technologies) and an ETDRS chart (McClenaghan et al, 2007); limits of agreement for test-retest repeatability were -0.13 logMAR to +0.17 logMAR for SSII and -0.12 logMAR to +0.13 logMAR for the ETDRS charts.

Blur-induced conditions: Testing with SS II under blur-induced condition showed no statistically significant difference between mean visual acuity on weeks 1 and 2 ($p>0.05$). However, the 95% limits of agreement for repeatability were equivalent to more than plus or minus 3 visual acuity lines (+0.330 to -0.321 logMAR). This result contrasts with some previously reported data for unaided visual acuity measurements (Lovie-Kitchen, 1988). In comparison to unaided visual acuity measurements, where subjects may have a large range of uncorrected refractive error, the present study tested adults with a myopic blur of -1.25 diopters over habitual correction. Kheterpal et al (Kheterpal et al, 1996) reported limits of

agreement between ±0.21 and ±0.25 logMAR (plus or minus 2 lines) for unaided visual acuity measurements, which is only slightly less than in the present study.

Statistical analysis of data in the present study demonstrated only a slight trend towards difference in test-retest limits of agreement for SS II vs. HOTV ($p = 0.19$). This is mostly due to the logMAR HOTV chart returning better limits of agreement ranging from +0.184 to -0.152 logMAR units (compared to +0.330 to -0.321 logMAR for SS II as mentioned above). However, the mean difference in visual acuity across sessions was actually smaller for SS II (+0.004) as opposed to HOTV (0.016), thus explaining the negative finding. Similar repeatability results in HOTV chart visual acuity were reported by Lovie-Kitchen (Kovie-Kitchen, 1988), with limits of agreement of ±0.16 logMAR for unaided visual acuity measurements.

Table 8: Previous studies showing repeatability of visual acuity measurements

Study	Participants	Method	SD* and 95% LOA*
Manny et al, 2003	N=86 mean age=9.4yo (6-11), normal myopic children	BCVA better than 0.2 logMAR, ETDRS chart x 2, 4 weeks apart	OD: SD ±0.075 OS: SD ±0.071 LOA ±0.15 both eyes
Chen et al, 1998	N=43 32 amblyopes mean age=6.4yo (3.6-11.9); 11 normal vision, mean age=6.7yo (4.8-8.6)	BCVA; Lea Symbol ETDRS-type chart, modified ETDRS fast protocol x 2, same session	Amblyopic eye SD ±0.33 normal eye SD ±0.16 LOA group mean ±0.18 LOA amblyopic eye ±0.18 LOA normal eye ±0.17
Kheterpal et al, 1996	N=18, mean age=4.8yo (3.5-8); refractive and binocular disorders, 3 with normal vision	Unaided VA, matching response logMar test x 2, 1 week apart	OD: LOA: ±0.21 OS: ±0.25 14 children with one eye consistently better better eye LOA: ±0.15 poorer eye LOA: ±0.22
Lovie-Kitchen, 1988	N=115 age range: 11-74yo, normal vision, students and patients	Unaided, Bailey-Lovie x 2, same session	SD ±0.08 LOA ±0.16
Elliott and Sheridan, 1988	N=20 mean age=64yo normal vision	BCVA and unaided Ferris logMar x 2, 1 week apart	Best refracted eye: LOA ±0.07 Unaided eye LOA ±0.21
Reeves at al, 1993	N=56 age range: 20-77yo, various clinical conditions with VA better than 20/60	BCVA, Bailey-Lovie x 2, 4 to 12 weeks apart	SD ±0.10 LOA ±0.19

Table 8: (continued)

Study	Participants	Method	SD* and 95% LOA*
Hazel et al, 2002	N=40 mean age= 28.4yo (±11.9) normal vision	Aided habitually corrected, ETDRS x 2, 1-2 weeks apart	SD ±0.10 LOA ±0.14
Present study (SS II)	N=30 mean age=24.6 yo (22-31) normal vision, students	Aided habitually corrected VA and myopic-blur induced VA logMAR HOTV x 2, 1-2 weeks apart	Habitual SD ±0.068 blur-induced ±0.189 Habitual LOA -0.152/+0.184 Blur-induced LOA -0.321/+0.330
Present study (EVA)	N=30 mean age= 24.6 yo (22-31) normal vision, students	Aided habitually corrected VA and myopic-blur induced VA logMAR HOTV x 2, 1-2 weeks apart	Habitual SD ±0.035 blur-induced ±0.158 Habitual LOA -0.050/+0.070 Blur-induced LOA -0.170/+0.163

* Reported in logMAR units

LOA = 95% limits of agreement; SD = standard deviation; BCVA = best-corrected visual acuity; VA = visual acuity; OD = right eye; OS = left eye

EVA: repeatability

Habitually-corrected conditions: No statistically significant difference was found between visual acuity tested with EVA on weeks 1 and 2 ($p>0.05$) under habitually-corrected conditions. The 95% limits of agreement ranged from +0.070 to -0.050 logMAR units, which is much smaller variability in visual acuity than previously reported (Table 8). Furthermore, the limits of agreement were significantly larger for the logMAR HOTV chart

(-0.160 / +0.162) compared to EVA ($p < 0.05$), thus suggesting poorer repeatability for the logMAR HOTV chart. In the present study EVA also performed superiorly when compared to previous reports that evaluated test-retest reliability of the Amblyopia Treatment Study (ATS) protocol presented on EVA (Holmes et al, 2001; Moke et al, 2001). Holmes at al found 95% limits of agreement to be ±0.125 logMAR for test-retest measurements, while Moke at al reported 95% limits of agreement to be ±0.13 logMAR. This represents approximately twice the variance in EVA VA measurements across testing sessions when compared to results in the present study (95% limits of agreement: +0.070 to -0.050, as mentioned above). However, both studies evaluated children ages from 2 to7 years old with unaided visual acuity, while the present study tested young adults with corrected refractive error. The lower limits of agreement in the present study might therefore be explained by more consistent performance (threshold criterion, attention, etc.) of adults compared to children.

Blur-induced conditions: Testing with EVA under blur-induced condition showed no statistically significant difference between mean visual acuity on weeks 1 and 2 ($p>0.05$). Similar to habitual-corrected condition, the limits of agreement were significantly larger for logMAR HOTV compared to EVA ($p < 0.05$). 95% limits of agreement ranged from +0.163 to -0.170 logMAR when tested with EVA vs. -0.316 to +0.384 logMAR when tested with HOTV. To date, there have been no studies evaluating EVA test-retest reliability separately under uncorrected versus aided visual acuity.

SS II: Validity

When testing with SS II, no statistically significant difference in visual acuity was found between SS II and logMAR HOTV charts for both habitually-corrected and blur-induced conditions. Under habitual correction, the mean difference in visual acuity between charts was -0.018 ± 0.063, with SS II showing a statistical trend towards underestimation of (poorer) visual acuity. The opposite results were reported by McClenaghan et al (McClenaghan et al, 2007) with mean visual acuity being 0.02 logMAR better on the SS II than ETDRS chart under best corrected visual acuity ($p < 0.05$). Individuals have been reported to consistently identify visual acuity changes only up to 0.025 logMAR (Miller et al, 1997). These mean differences are very small (0.02 logMAR is equivalent to 1 optotype on a line of 5 optotypes) and are unlikely to be meaningful in clinical practice. The 95% limits of agreement for SS II vs. logMAR HOTV ranged from +0.104 to -0.141 logMAR under habitually-corrected condition and from +0.353 to -0.345 logMAR under blur-induced condition, thus demonstrating higher variability in visual acuities under blur-induced conditions.

EVA: Validity

A small but statistically significant difference ($p<0.05$) was found between EVA and logMAR HOTV chart under habitually-corrected condition. EVA underestimated HOTV visual acuity by a mean -0.087 logMAR (approximately 1 line on logMAR chart). Given the previously discussed large test-retest fluctuations in VA estimation (consistent with reports in the literature), these findings regarding EVA validity might not necessarily be of immediate

clinical relevance. Alternatively, any clinically significant difference between EVA and logMAR HOTV chart VA estimations may not necessarily represent a symptom of limited repeatability: rather, it could be the consequence of the inherent testing procedure being implemented by the EVA system, where the lowest reachable threshold for visual acuity is 20/16 (as opposed to 20/10 when tested with logMAR HOTV chart). Finally, the differences in threshold definition (75% correct for the automated ATS system used by EVA as opposed to the 60%-correct criterion for logMAR HOTV) might be responsible for these findings.

Under blur-induced conditions, no significant difference in visual acuity estimation was found between EVA and HOTV charts under blur-induced condition ($p=0.89$). Limits of agreement between EVA and HOTV ranged from +0.056 to -0.230 logMAR in habitually-corrected conditions and from +0.246 to -0.253 logMAR under blur-induced condition.

Habitually-corrected vs. Blur-induced Conditions

Based on the results of this study all charts (standard logMAR HOTV chart and test devices, SS II and EVA) demonstrated almost double the variation in visual acuity under 1.25D myopic blur-induced conditions. These results support previous findings of wider limits of agreement for unaided visual acuity as opposed to aided visual acuity (Table 8).

Carkeet et al (Carkeet et al, 2001) reported a larger probit size* (up to 0.12 logMAR) with optical defocus induced by +1.00D and +2.00D lenses. In all instruments, including the logMAR HOTV standard chart, the distribution of visual acuity measurements was on average two times as wide under blur-induced conditions. This corresponds to the blur-induced condition resulting in twice the variability in VA estimations compared to the

habitual-corrected condition. These differences could be of importance in situations when visual acuities are compared between multiple visits if the patient is not wearing appropriate corrective lenses. Therefore, larger variation of visual acuity should be expected for each visual acuity device used in vision screening in young adults and children who have uncorrected myopic refractive error.

Conclusions

In summary, both computerized systems tested in the present study (SS II and EVA) measure VA as reliably as the logMAR HOTV chart, i.e. with adequate repeatability and validity (compared to logMAR HOTV chart). These instruments are therefore good devices to test crowded HOTV visual acuity in young healthy adults. The SS II device has an advantage in terms of portability. Results from the EVA device showed superior test-retest reliability, suggesting improved precision compared to SSII and logMAR HOTV, but also points towards underestimation of VA by the EVA system under habitually-corrected conditions. Larger variability in visual acuity measurements should be expected under myopic blur conditions, and potentially also when the instruments are used for screening purposes.

* Probit size is defined as the standard deviation of the cumulative normal distribution of visual acuity, and mathematically describes how gradually the subject(s) experiences transition from seeing to non-seeing. Larger probit size corresponds to a flatter slope of the psychometric function, suggesting larger variability between visual acuity measurements

Future studies will be required to explore the validity and test-retest reliability of SS II and EVA (as opposed to the logMAR chart) in specific patient groups and/or clinical scenarios. Comparing performances of computerized devices to that of the HOTV chart in and adequately large sample size of children will provide further indications about the generalizability of results presented in this study. Similarly, patients with different moderate to high refractive error other than myopic blur (e.g. hyperopia, astigmatism) should be included in future studies to provide estimates of repeatability for each VA device. These analyses will provide further insight into the value of SS II and EVA for screening purposes. Finally, larger studies enrolling both adults and children and including the ETDRS Sloan Letters chart among tested devices will provide more accurate and generalizable estimates of limits of agreement for VA estimation (both test-retest and across devices), that will inform clinical practice and conduct of future research studies.

BIBLIOGRAPHY

Altman DG, Bland JM, 1983. Measurement in medicine: the analysis of method comparison studies. Statistician. 32, 307–17

Arditi A, Cagenello R, 1998. On the statistical reliability of letter-chart visual acuity measurements. Invest Ophthalmol Vis Sci. 34, 120-29

Bailey IL, Lovie JE, 1976. New design principles for visual acuity letter charts. Am J Optom Physiol Opt. 53, 740-45

Bland JM, Altman DG, 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1, 307–10

Brown B, Lovie-Kitchin J, 1993. Repeated visual acuity measurements: establishing the patient's own criterion for change. Optom Vis Sci. 70, 45-53

Campbell FW, GreenD, 1965. Optical and retinal factors affecting visual resolution. J Physiol. 181, 576-93

Carkeet A, 2001. Modeling logMAR visual acuity scores: effects of termination rules and alternative forced-choice options. Optom Vis Sci. 78, 529-38

Carkeet A, Lee L, Kerr JR, et al, 2001. The slope of the psychometric function for Bailey-Lovie letter charts: defocus effects and implications for modeling letter-by-letter scores. Optom Vis Sci. 78, 113-21

Chen SI, Chandna A, Norchia AM, et al, 2006. The repeatability of best corrected acuity in normal and amblyopic children 4 to 12 years of age. Invest Ophthalmol Vis Sci. 47, 614-19

Cotter SA, Tarczy-Hornoch K, Wang Y et al; Multi-Ethnic Pediatric Eye Disease Study Group., 2007.Visual acuity testability in African-American and Hispanic children: the multi-ethnic pediatric eye disease study. Am J Ophthalmol. 144, 663-7

Cryotherapy for Retinopathy of Prematurity Cooperative Group, 1996. Multicenter trial of cryotherapy for retinopathy of prematurity: Snellen visual acuity and structural outcome at 5½ years after randomization. Arch Opthalmol. 114, 417-24

De Young-Smith M, Baker J, 1986. A comparative study of visual acuity tests. Am Orthoptic J. 36, 160-4

Early treatment Diabetic Retinopathy Study: Manual of Operations. Baltimore, ETDRS Coordinating Center, Department of Epidemiology and preventive medicine, 1980; ch 12: 1-15

Elliott DB, Sheridan, 1988. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. Ophthalmic Physiol Opt. 8, 397-401

Ferris III FL, Bailey I, 1996. Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the eye care technology forum. Ophthalmology. 103, 181-2

Ferris FL III, Kassoff A, Bresnick GH, et al, 1982. New visual acuity charts for clinical research. Am J Ophthalmol. 94, 91-6

Ferris III FL, Sperduto RD, 1982. Standardized illumination for Visual Acuity testing in clinical research. Am J Ophthalmol. 94, 97-8

Flom M, Weymouth W, Kahneman D, 1963. Visual resolution and contour interaction. J Opt Soc Am. 53, 1026-32

Friendly DS, 1978. Preschool visual acuity screening tests. Trans Am Ophthalmol Soc. 76, 383-480

Hazel CA, Elliott BD, 2002. The dependency of logMar visual acuity measurements on chart design and scoring rule. Optom and Vis Sci. 79, 788-92

Hered R, Murthy S, Clancy M, 1997. Comparison of the HOTV and Lea symbols charts for preschool vision screening. J Pediatr Ophthalmol Strabismus. 34, 24-8

Holmes J, Beck R, Repka M, et al, 2001. The Amblyopia treatment study visual acuity testing protocol. *Arch Ophthalmol*. 119, 1345-53

Johnson CA, Casson EJ, 1995. Effects of luminance, contrast and blur on visual acuity. Optom Vis Sci. 72, 864-9

Kheterpal S, Jones HS, Auld R, et al, 1996. Reliability of visual acuity in children with reduced vision. Ophthalmic Physiol Opt. 16, 447-9

Kulp MT, Dobson V, Peskin E et al, 2004. Vision in Preschoolers Study Group. The electronic visual acuity tester: testability in preschool children. Optom Vis Sci. 81, 238-44

Lippmann O, 1969. Vision of young children. Arch Ophthalmol. 81, 763-775

Lovie-Kitchen JE, 1988. Validity and reliability of visual acuity measurements. Ophthalmic Physiol Opt. 8, 363-70

Manny RE, Hussein M, Gwiazda J, et al, 2003. Repeatability of ETDRS visual acuity in children. Invest Ophthalmol Vis Sci. 44, 3294-3300

Mayer DL, Gross RD, 1990. Modified Allen pictures to assess Amblyopia in young children. Ophthalmology. 97, 827-832

McClenaghan N, Kimura A, Stark LR, 2007. An evaluation of the M&S technologies Smart System II for visual acuity measurement in young visually-normal adults. Optom Vis Sci. 84, 218-23

Merritt J, Game S, Williams D, et al, 1996. Visual acuity in preschool children: the Chapel Hill-Durham day-care vision study. J Natl Med Assoc. 88, 709-12

Miller AD, Kris MJ, Griffiths AC, 1997. Effect of small focal errors on vision. Optom Vis Sci. 70, 45-53

Moke PS, Turpin AH, Beck RW, et al, 2001. Computerized method of visual acuity testing: adaptation of the Amblyopia treatment study visual acuity testing protocol. Am J Ophthalmol. 132, 903-9

Muller P, 1951. The vision of amblyopes. Ophthalmologica. 121, 143-9

National Research Council (NRC), 2002. Visual impairments – Determining eligibility for social security benefits, first ed. National Academy Press, Washington, DC.

Neuhäuser M, Manly BF, 2004. The Fisher-Pitman permutation test when testing for differences in mean and variance. Psychol Rep. 94, 189-94

Raasch TW, Bailey IL, Bullimore MA, 1998. Repeatability of visual acuity measurement. *Optom Vis Sci*. 75, 342-48

Reeves BC, Wood JM, Hill AR, et al, 1993. Reliability of high- and low-contrast letter charts. Ophthalmic Physiol Opt. 13, 17-26

Rice ML, Leske DA, Holmes JM, 2004. Comparison of the amblyopia treatment study HOTV and electronic-early treatment of diabetic retinopathy study visual acuity protocols in children aged 5 to 12 years. Am J Ophthalmol. 137, 278-82

Sheridan MD, 1960. Vision screening of very young and handicapped children. BMJ. 2, 453-6

Simpson TL, Barbeito R, Bedell HE, 1986. Effect of optical blur on visual acuity for targets of different luminances. Ophthal Physiol Opt. 6, 279-81

Stuart JA, Burian HM, 1962. A study of separation difficulty. It's relationship to visual acuity in normal and amblyopic eyes. Am J Ophthalmol. 53, 471-7

Thorn, F, 1990. Effects of dioptric blur on Vistech contrast sensitivity test. Optom Vis Sci. 67, 8-12

Vision in preschooler (VIP) study group, 2003. Threshold visual acuity testing of preschool children using the crowded HOTV and Lea symbols acuity tests. J AAPOS. 7, 396-9

Vision in preschooler (VIP) study group, 2004a. The Electronic Visual Acuity Tester: Testability in preschool children. Optom Vis Sci. 81, 238-44

Vision in preschooler (VIP) study group, 2004b. Preschool visual acuity screening with HOTV and Lea Symbols: testability and between-test agreement. Optom Vis Sci. 81, 678-83