

The Repeatability of Best Corrected Acuity in Normal and Amblyopic Children 4 to 12 Years of Age

Sean I. Chen,^{1,2,3} Arvind Chandna,¹ Anthony M. Norcia,³ Mark Pettet,³ and Deborah Stone¹

PURPOSE. The main purpose of this work was to measure repeatability of line-by-line logMAR (logarithm of the minimum angle of resolution) acuity in normal and amblyopic children, while adequately controlling for optical defocus.

METHODS. The Lea Symbols Chart is a constant-crowding, equal-logMAR increment chart similar in design to the Early Treatment Diabetes Retinopathy Study [ETDRS] chart. LogMAR visual acuity was tested twice in each eye of 32 amblyopic and 11 normal children. Each test commenced with screening in which one of the three central symbols was chosen for identification starting with the 1.0- or 0.9-logMAR line, progressing to every second line until incorrect identification occurred. Symbol-by-symbol presentation then commenced at the logMAR line containing the last correctly identified symbol. The threshold was recorded as the last logMAR line where four of four or four of five correct responses occurred (i.e., line-by-line scoring). Retesting by the same examiner was identical and occurred within the same session.

RESULTS. There was no significant difference in repeatability among normal, fellow, or amblyopic eyes. The difference between test and retest thresholds lay between ± 0.10 logMAR in 93% of eyes. The 95% limits of agreement for the difference was ± 0.18 logMAR. Repeatability in eyes tested first did not differ from that in those tested second in either the normal or amblyopic groups.

CONCLUSIONS. In the age-group tested, the line-by-line method of threshold scoring compares favorably with previous reports of both line-by-line and interpolated threshold scoring. There was no clinically meaningful difference in repeatability between the normal and amblyopic children tested. (*Invest Ophthalmol Vis Sci.* 2006;47:614–619) DOI:10.1167/iovs.05-0610

Knowledge of an instrument's measurement variability (measurement noise) is an essential aspect of separating normal from abnormal states. It is also important in the detection of significant change in the longitudinal assessment of any clinical condition. Test-retest reliability is a measure of a test's precision, rather than its accuracy, since the latter entails how closely matched the measurements are to their true value.

From the ¹Vision Assessment Unit, Department of Ophthalmology, Royal Liverpool Children's Hospital, Liverpool, United Kingdom; the ²Unit of Ophthalmology, University of Liverpool, Liverpool, United Kingdom; and the ³Infant Vision Laboratory, The Smith-Kettlewell Eye Research Institute, San Francisco, California.

Supported by Guide Dogs for the Blind Association Project Grant OR2001-99a and The Smith Kettlewell Eye Research Institute.

Submitted for publication May 17, 2005; revised July 25, September 13, and October 25, 2005; December 15, 2005.

Disclosure: **S.I. Chen,** None; **A. Chandna,** None; **A.M. Norcia,** None; **M. Pettet,** None; **D. Stone,** None

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Sean I. Chen, Vision Assessment Unit, Department of Ophthalmology, Royal Liverpool Children's Hospital, Eaton Road, Liverpool L12 2AP, UK; s.chen@blueyonder.co.uk.

Regarding the precision of a visual acuity test, test-retest variability, interobserver and intraobserver error, and focusing errors are among the main contributors to measurement variability. The largest-magnitude measurement error determines the precision of any instrument. Previous work has demonstrated that when the test procedure is rigidly controlled, observer errors can be small.^{1,2}

Visual acuity is an important clinical measure of visual function. The ETDRS-type chart and scoring protocol³ standardizes the administration and scoring of visual acuity and has arguably become the gold-standard instrument in both research and clinical settings. The method of threshold scoring used in the ETDRS protocol is that of letter-by-letter, or interpolation. In this method, each letter, or optotype, presented is assigned an equal logMAR score. Charts displaying five optotypes per line and graduated in 0.10-logMAR increments therefore assign a 0.02-logMAR score per optotype.

An alternative method of determining acuity thresholds is the line-by-line method. This method of assessment interprets each line as an individual psychophysical task that the patient must pass to progress. The line-by-line scoring system is accepted by authoritative bodies^{4,5} and is also in widespread use in clinical and pediatric research settings,^{6–11} including The Amblyopia Treatment Study.¹

The stated benefit of interpolation (letter-by-letter) is that it is theoretically more sensitive to the measurement of change owing to the subdivision of the 0.10 logMAR incremental graduations of each chart line into an incremental scale five times finer (each optotype becomes an increment).^{12–14} The implication is that the line-by-line scoring rules may be insufficiently sensitive to detect and monitor treatment effects in for example amblyopia. However, a recent modeling study¹⁵ has shown relatively small differences between the two termination rules.

Empiric limits of the line-by-line and interpolation rules have been established only in adults and may differ in pediatric populations. Furthermore, despite a theoretical advantage in favor of the interpolation method, Carkeet¹⁵ makes a practical recommendation of using a line-by-line termination rule.

In addition to determining the test-retest reliability of line-by-line scoring in children, we also wanted to control for the effects of uncorrected refractive error. Several studies^{16–18} have indicated that uncorrected refractive error contributes to test-retest variability. A recent study involving normal adults¹⁸ highlighted wide variations in published repeatability data. Based on their work and that of others,^{6,12–14,16–22} Rosser et al.¹⁸ have suggested (hyperopic) optical defocus as one of the main sources of repeatability error. Other sources of error, such as the presence or absence of disease,^{14,16,21} whether the measurements were conducted in a single session,^{6,16} and the subject's age^{19,20} have been considered and found not to contribute significantly to repeatability error. Of note regarding the issue of subject age, however, Beck et al.¹⁹ tested mostly adults with a mean age of 50 ± 22 years (SD) and Lovie-Kitchin and Brown²⁰ tested adults older than 20 years.

Many studies^{1,2,8–11,23–27} deal with repeatability of recognition acuity in the <12-year-old group. In those in which crowd-

ing control was uniformly used^{1,2,9,10,23-26} and amblyopes and normal subjects were compared,^{1,2,23,24,26} only Holmes et al.¹ and Moke et al.² used line-by-line scoring. Furthermore, regardless of the method of scoring, only Moke et al.² controlled for optical defocus over most (87.5%) of their study population. The test used by Moke et al. is a computer-based version of the HOTV test, which is not commercially available and costs approximately US \$750 (raw materials and customization costs). In addition, the majority (78%) of their amblyopic group had acuities better than or equal to 0.40 logMAR (20/50).

The designs of previous studies in this area (i.e., repeatability of visual acuity in children) have ignored amblyopes, not uniformly controlled for optical defocus, or used acuity testing devices that are not commercially available. In this study, we addressed all the limitations of the prior studies. We chose crowded Lea symbols because, compared with HOTV, it has been shown to have superior testability in two^{10,28} of three^{10,28,29} large-scale pediatric studies. In addition, it is commercially available, inexpensive, and portable; is in widespread use; and most important, has been standardized.³⁰ We chose the line-by-line scoring method because it also is in widespread use.^{1,6-11}

In summary therefore, the primary purpose of this study was to inform clinical procedure by investigating test-retest variability of line-by-line scoring of the Lea chart in normal and amblyopic children having corrected uniformly for optical defocus in a standard protocol.

METHODS

Patients and Subjects

From November 2001 to September 2003, 71 children with diagnosed amblyopia and 21 children acting as normal control subjects were examined as part of a longitudinal study of amblyopia treatment. All patients were recruited consecutively from referral letters to one of the authors (AC). After full ophthalmic examination and cycloplegic refraction, patients were allowed to become accustomed to their spectacles.

Children prescribed spectacles were assessed every 4 weeks, to determine whether they were as yet accustomed to them. In amblyopes, this was deemed to have occurred on the first occasion after the spectacle prescription was issued when the fellow eye's acuity became equal to or better than its previously recorded threshold in the unaided state (mean duration, 7.8 ± 3.4 weeks [SD]). In normal children, this criterion was deemed to have been met when the acuity in both eyes achieved the threshold recorded in the unaided state or better, which required 8.4 weeks on average (with some noncompliance, $n = 2$). At this time, the child was reexamined, and the diagnosis of amblyopia or normal control was made.

Amblyopia was diagnosed when the interocular acuity difference (IAD), measured with a 3-m Crowded Lea Symbol chart (catalog no. 2503; Precision Vision, Bloomington, IL), was ≥ 0.20 logMAR. Normal subjects were recruited from referrals for suspected amblyopia that was disproved after identical examination, and also through local newspaper announcements. All normal participants were required to have been born fewer than 15 days prematurely, to have had a history of normal birth weight and, to be free of systemic disease. Furthermore, undilated ophthalmic examination by a fully qualified ophthalmologist (SC, AC) confirmed no ocular disease in addition to bifoveal fixation (4-D base-out test, 3/4 criterion) and, in those able to perform it, normal stereothreshold (Frisby near stereotest: two-up, one-down staircase). Potentially normal recruits found to have greater than 1 D of sphere (hyperopia or myopia) and/or greater than 0.5 D of cylinder (hyperopia or myopia) on manifest retinoscopy ($n = 2$) were subjected to cycloplegic refraction, corrected with full spectacle prescription, and confirmed to surpass normality criteria before recruitment. Only

data collected after participants became accustomed to their spectacles were included in the analysis.

The entire study group consisted of 32 amblyopes of mean age 6.44 years (range, 3.63-11.90 \pm 1.94; anisometric, 15; strabismic, 6; anisostabismic, 11) and 11 normal subjects of mean age 6.69 years (range, 4.84-8.58 \pm 1.12). Slightly more difficulty was encountered in recruiting younger normal subjects, and although there was a difference in age range between the two groups, it did not reach statistical significance ($P = 0.69$). Mean logMAR acuity of the amblyopic eyes was 0.50 ± 0.31 (SD), of the fellow eyes was 0.13 ± 0.15 , of the normal right eye was 0.10 ± 0.09 and, of the normal left eye was 0.06 ± 0.13 . The mean spherical equivalent refraction of the amblyopic eye was $+4.66 \pm 3.44$ D (SD) and of the fellow eye was $+2.83 \pm 2.42$ D. Only two children in the normal group needed spectacles (mean spherical equivalent right, $+4.94$ D, and left, $+4.63$ D) but each achieved normal acuity (mean logMAR of 0.08 in each eye) and stereoacuity (20 seconds of arc) for their age.

Local IRB approval was received. The protocol was in compliance with the tenets of the Declaration of Helsinki, and written, informed consent was obtained.

Study Design

Only participants providing complete data for the test and retest procedure in both eyes contributed data to the study. Data were collected by one author (SC) at each clinic session during the course of a controlled longitudinal study of amblyopia treatment. Most of the participants had more than one clinic session in which test and retest thresholds were collected in each eye, and therefore, in these instances, one record was randomly (computer random number generator) chosen in each patient for analysis. The visual acuity test (first determined threshold for a given eye) and retest (second determined threshold for that eye) procedures were identical and adhered rigidly to the protocol that will be described in the Procedure section. This procedure was based on the ETDRS-Fast method.³¹ For the younger age group of our participants a modification was necessary to shorten the testing procedure. Monocular acuity thresholds were determined for each eye during the test phase. Each eye was then retested (not necessarily in the same sequence) within a 20-minute period. The threshold determinations for each eye were interleaved with orthoptic testing and breaks as necessary to maintain the interest and cooperation of the child.

At each visit, the acuity was attempted first as a matter of course and without knowledge of previous acuity thresholds. For previously noted reasons³² and reasons related to pediatric cooperation levels, there was limited ability to mask the tester from the test thresholds when conducting the retest thresholds. The tester was fully aware, however, of possible bias in this respect and adhered rigidly to the acuity testing procedure at all times (see the Discussion section for further detail).

Procedure

Best corrected distance acuity was determined with the same ETDRS-type, high-contrast, standardized, retroilluminated chart in the same room throughout. The mean luminance was 266 cd/m², in compliance with recommendations for the standardization of acuity measurement.^{4,33} The chart is logarithmically calibrated and, at the specified testing distance, ranges in logMAR from 1.0 to -0.4 . The interoptotype distances relate to the size of the symbols, of which there are four: circle, house, heart, and square. The symbols are chosen to be universally recognizable among children, and each differs in only a few details from a circle, so that, beyond the threshold of recognition, all shapes blur to resemble a circle.³⁴ The procedure for test and retest threshold determination was identical and involved recording a threshold determined by line-by-line rather than interpolated³⁵ scoring. As mentioned earlier, line-by-line scoring interprets each line as an individual psychophysical task that the patient must pass to progress. The four-alternative choice design of the Lea chart in this study dictated

TABLE 1. Limits of Agreement for Line-by-Line ETDRS-Type Crowded Lea Symbol Acuity in Children under 12 Years of Age

Participant Group	Participants (n)	Eye	Eyes (n)	Mean Acuity Difference*	SD of Acuity Difference*	95% LOA*
Amblyopes	32	Amblyopic	32	0.00	0.09	0.17
		Fellow	32	-0.01	0.09	0.18
Normal Subjects	11	Right	11	0.01	0.09	0.19
		Left	11	-0.05	0.07	0.14
Entire group	43	Amblyopic/left	43	-0.01	0.09	0.18
		Fellow/right	43	-0.01	0.09	0.18

Data are expressed in LogMAR units. Acuity difference, test threshold - retest threshold. 95% LOA = $2 \times$ SD of acuity difference. The 95% LOA for the right eye of normal subjects represents the rounded (2 decimal places) result of two times the calculated standard deviation and not the result of the direct product of 2×0.09 . No significant difference in repeatability between amblyopic and normal observers was found, so data were pooled for the entire group.

that the subject surpass 62.5% (halfway between chance and 100%) correct responses for each size optotype (logMAR line) before progressing. Because three of five represents 60% correct, in this study a four of five (80% correct) criterion or better was universal.

The first stage of testing involved a screening stage in which the child was requested to identify one of the central three symbols chosen randomly by the tester, starting from the 1.0 or the 0.9 logMAR line (if the 0.9 line was incorrectly identified, the 1.0 symbol was used). No second chances were given at any time, and the child was expected to guess if uncertain. If the child was unable to identify correctly the first symbol presented from the 1.0-logMAR line, the testing distance was reduced in 1-m increments to a minimum of 1 m from the chart and then to 0.5 m from the chart. Screening progressed in -0.2-logMAR increments (every second line), avoiding the presentation of circles, until an incorrect response was recorded. At this point, screening stopped, and the line containing the previously correct response was tested from left to right or vice versa. A threshold was recorded as the logMAR of the line last correctly (4/4 or 4/5 correct) identified (i.e., the line just above the one missed).

Several measures were used to achieve a reliable threshold in cases in which false starts (child unwilling to cooperate with identifying symbols e.g., due to playfulness or fussiness) were encountered. For example, the design of this chart (catalog no. 2503; Precision Vision) provides space for two additional sets of symbols on either side of the central set (below the 0.50-logMAR level), allowing the tester to obviate any attempts by the child to memorize the symbols. If false starts occurred before the 0.50-logMAR level, the child's view of the chart was shielded, and the child was misinformed that the entire chart was being changed, a break would be given, and testing would resume when cooperation was again appropriate.

Statistical Analysis

All units of acuity were recorded and analyzed in logMAR notation. The 95% limits of agreement (LOA) as described by Bland and Altman³² were evaluated for the entire group and for each subgroup. Student's *t*-test was used to evaluate the differences (difference = test threshold - retest threshold) between eyes and also in groups of participants, to evaluate whether differences were affected by the sequence of testing (i.e., results of first eye chosen to be tested compared with that of the second eye chosen to be tested).

RESULTS

Considering the group of eyes first tested in the test phase of data collection (24 amblyopic, 8 fellow, and 11 normal eyes), there was a statistically significant difference ($P = 0.0002$) in mean acuity thresholds (average of test and retest threshold) between amblyopic (0.52 ± 0.33 logMAR) and normal (0.08 ± 0.09 logMAR) as well as amblyopic and fellow (0.18 ± 0.17

logMAR) eyes ($P = 0.01$). There was no significant difference between fellow and normal eyes ($P = 0.12$) in this group. Concerning IAD within this group, there was no difference ($P = 0.39$) in the degree of amblyopia between amblyopes who had their amblyopic eye selected for testing first ($n = 24$; mean IAD, 0.40 ± 0.37 logMAR) compared with their fellow eye selected for testing first ($n = 8$; mean IAD, 0.29 ± 0.16 logMAR). No effect on repeatability (magnitude of acuity differences between test and retest thresholds) was found related to the sequence of testing (first versus second) either for normal subjects ($P = 0.14$) or amblyopes ($P = 0.68$). Also, irrespective of which eye was selected (first or second), there was no significant difference in repeatability between normal and amblyopic observers.

The line-by-line method categorizes repeatability in 0.10 logMAR increments starting from 0.00 (no variation). Considering the lack of significant difference in repeatability related either to the sequence of testing or the diagnostic category, the repeatability for the entire group (normal and amblyope) was calculated. A difference of ± 0.10 logMAR between test and retest thresholds occurred in 93% of cases, and a difference of ± 0.20 logMAR occurred in 7% of cases. No significant difference ($P = 0.13$) was found between the magnitude of threshold improvement (26/48) versus disimprovement (22/48) on retesting.

In the entire study group, the mean difference between the test threshold and that of the retest was an improvement in threshold of 0.01 logMAR (Table 1), and the largest difference that occurred was an improvement of 0.05 logMAR (2.5 symbols), neither of which is clinically meaningful.

The 95% LOA³² are also reported for the purpose of comparison with previous reports, and those for the entire group and subgroups are provided in Table 1. The mean versus difference plots of test-retest thresholds for the normal group, the patients with amblyopia, and the entire group are shown in Figures 1A, 1B, and 1C, respectively. The dashed lines represent 93% of the study group. In all cases, the difference between test and retest threshold lay between ± 0.20 logMAR. Regardless of mean acuity (Fig. 1C), the distribution in the entire group was clustered largely between ± 0.10 logMAR (dashed lines, 93%). The range of acuity thresholds for the entire group was -0.10 to 1.50 logMAR (mean, 0.26 ± 0.29). Finally, test-retest reliability was similar over the range tested, as can be seen in Figures 1A, 1B, and 1C.

DISCUSSION

Crowded logMAR acuity is arguably the gold-standard visual acuity measure in adults and is rapidly becoming so in children, especially with the recalibration (www.lea-test.fi/; available at no charge) of the Lea symbol chart¹⁰ to the international

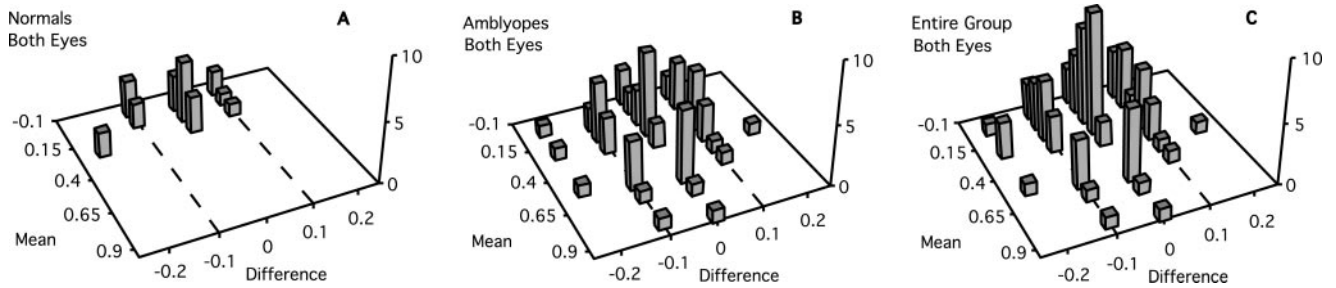


FIGURE 1. Mean versus difference of test-retest acuity in 22 eyes of normal children (A), in 64 eyes of amblyopic children (B) and, in 86 eyes of the entire study group (C). In most instances (93%) repeatability lies within ± 0.10 logMAR, delineated by *dashed lines*. Repeatability was not overtly dependent on mean acuity (range, -0.10 to $+1.50$ logMAR).

standard, the Landolt-C chart.^{4,5} Although advances in standardizing the optotype acuity measurement procedure have been made in adults,^{3,4,35} these have been comparatively few in the pediatric age group. Repeatability statistics for crowded logMAR acuity in the pediatric age group have been reported in some studies,^{1,2,8-11,23-27} but in only two of these^{2,25} was optical defocus controlled. Manny et al.²⁵ assessed only normal myopic children (Table 2) and Moke et al.² included amblyopes with acuities ≤ 0.4 logMAR. In the present study and that of Moke et al.,² both normal subjects and amblyopes were assessed, with no difference found in repeatability between these two groups. In addition, we and Moke et al.² (as well as Holmes et al.,¹ in which optical defocus was not reported as uniformly controlled¹), found that, in 93% of cases, test and retest thresholds lay within ± 0.10 logMAR. This study and the

other two (Moke et al.² and Holmes et al.¹) used line-by-line scoring.

The line-by-line scoring system is favored by clinicians^{6,15} and approved by authoritative bodies.^{4,5} The system used in this study would necessarily increase the 95% LOA,¹²⁻¹⁴ but our results and those of others^{1,2} indicate that the method is capable of detecting acuity differences of greater than 0.10 logMAR (± 5 symbols). If constrained to reporting the 95% LOA, however, the repeatability increases to ± 0.18 logMAR on average (maximum, ± 0.19) as predicted¹²⁻¹⁴ but still not dissimilar to most other reports,^{1,2,25} including those using interpolated scoring (Table 2). This is consistent with modeling experiments comparing these two types of termination rules that suggest that differences between the two are small.¹⁵

TABLE 2. Repeatability of Acuity in Normal and Amblyopic Children under 12 Years of Age

Study	Participants	Method	95% LOA
Kheterpal et al. ²³	$n = 18$; mean age: 4.8 y (3.5-8); three normal subjects	Interpolated scoring; unaided monocular logMAR VA $\times 2$, 1 week apart	OD: ± 0.21 logMAR; OS: ± 0.25 logMAR; 14 children with one eye consistently better; better eye: ± 0.15 logMAR; worse eye: ± 0.22 logMAR
McGraw et al. ²⁴	$n = 68$; mean age: approximately 5.32 y; normal subjects	Interpolated scoring; unaided monocular VA $\times 2$, 4 weeks apart; GAC logMAR acuity (line acuity with crowding box surround)	± 0.10 LogMAR (± 4 letters)
Manny et al. ²⁵	$n = 86$; mean age: 9.4 y (6-11); normal myopes (baseline VA better than 0.2 LogMAR)	Interpolated scoring; BCVA, ETDRS chart $\times 2$, 4 weeks apart	± 0.15 LogMAR (± 8 letters)
Holmes et al. ¹	$n = 88$; mean age: 5.3 y (3-7); 59 normals, 25 amblyopes, 4 developmental delay	Line-by-line scoring; nonamblyope refraction unspecified; isolated, surrounded HOTV (BVAT) $\times 2$, same session	Group mean: ± 0.18 LogMAR (± 9 letters)
Moke et al. ²	$n = 156$; mean age: 5.6 y (3-7); 76 normals subjects, 60 amblyopes, 20 abnormal subjects	Line-by-line scoring; BCVA in 87.5%; isolated, surrounded HOTV (custom-built apparatus) $\times 2$; same session	Group mean: ± 0.19 LogMAR (± 9.5 letters)
Present study	$n = 43$ of 92 in longitudinal amblyopia treatment study; 32 amblyopes, mean age: 6.4 y (3.6-11.9); 11 normals, mean age 6.7 y (4.8-8.6)	Line-by-line scoring; BCVA; Lea Symbol ETDRS-type chart; modified ETDRS-Fast (Camparini ³¹) protocol $\times 2$; same session	Group mean: ± 0.18 LogMAR (± 9 symbols); amblyopes: ± 0.18 LogMAR; normal subjects: ± 0.17 LogMAR

In practical terms, when using line-by-line scoring, however, a 95% LOA of ± 0.18 would effectively amount to a threshold for detecting a real difference of 2 lines, and in our sample 100% of test and retest scores lay within ± 0.2 logMAR—generally the accepted threshold for the detection of amblyopia and better than the interocular difference threshold specified in a recent large amblyopia study.³⁶

Despite their use of the interpolated scoring method, the reported 95% LOA in Khetarpal et al.²³ are actually greater than those reported by us (Table 2). This most likely reflects the smaller study group and possibly also the effect of optical defocus.¹⁶⁻¹⁸ Whether the careful control of optical defocus affected our results and those of Moke et al.² is unclear, since Holmes et al.¹ reported similar results in a group of children in which the refractive status of the normal subjects was not specified. Unfortunately, recalculation of our data using the interpolated scoring system is not possible, because the study criteria recorded only line-by-line thresholds from the outset.

Our analysis failed to identify any statistically significant difference in the magnitude of threshold differences between test and retest of the normal and the amblyopic groups. This result is in keeping with those in previous reports in both adult and pediatric groups, which showed either no statistically significant difference in repeatability^{1,2,14,21} between normal and abnormal groups or, if a difference was found (± 0.02 logMAR greater test-retest variability in cataract patients than normal),¹⁶ it was not clinically meaningful.

The adequacy of masking in the single observer, single-session data collection design of this study is a possible source of bias. In young pediatric age groups, cooperation is notoriously fickle, and in this study, intervals between acuity thresholds were occupied by conducting other orthoptic assessments rather than allowing the child to leave the examination room for non-clinic-related activity. It was decided also to collect test and retest data within the same session, because there is a high nonattendance rate (24.76% in the research clinic, 19.66% in the nonresearch clinic), thereby reducing the total number of visits required. Despite this, no significant difference was detected between the magnitude of improvement versus disimprovement ($P = 0.13$) for the entire group of eyes in which a change in threshold occurred ($n = 48$). A further argument against the possibility of bias in this report is related to a crucial finding in the work (i.e., the magnitude of repeatability is similar across three classes of eyes: normal, fellow, amblyopic). To have achieved this result, one would have had to inflate or deflate or adjust the bias systematically, depending on the group being tested at the time. We did not make such adjustments.

The Lea symbol chart has been shown to have testability superior to that of the HOTV test,^{10,28} but the question of its reliability, particularly in amblyopic children and also under conditions of appropriate optical correction has been investigated in this study. We have now shown that the Lea symbol chart, scored using a line-by-line method, produces test-retest reliability that is comparable to the HOTV test scored by the same method,^{1,2} and reliability that is comparable to other reports^{23,25} in which interpolated scoring was used. Thus, a degree of accuracy comparable to that of the HOTV design can be obtained with a test that has been shown to have superior testability in children. Best corrected acuity scores were measured throughout this study, and therefore the effect of optical defocus on repeatability was not measured, but there is no reason to suspect that it would be any less than that reported by Rosser et al.¹⁸ in normal adults. There appeared to be no clinically meaningful difference in repeatability between the normal and amblyopic children tested in this study.

References

- Holmes JM, Beck RW, Repka MX, et al. The amblyopia treatment study visual acuity testing protocol. *Arch Ophthalmol*. 2001;119:1345-1353.
- Moke PS, Turpin AH, Beck RW, et al. Computerized method of visual acuity testing: adaptation of the amblyopia treatment study visual acuity testing protocol. *Am J Ophthalmol*. 2001;132:903-909.
- ETDRS. Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy Study report number 1. Early Treatment Diabetic Retinopathy Study research group. *Arch Ophthalmol*. 1985;103:1796-1806.
- Colenbrander A. Visual Standards aspects and ranges of vision loss with emphasis on population surveys. *Report for the International Council of Ophthalmology*. 2002;:1-33.
- National Research Council. Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of Working Group 39, Committee on Vision. Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, DC. *Adv Ophthalmol*. 1980;41:103-148.
- Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci*. 1993;34:120-129.
- Becker R, Hubsch S, Graf MH, Kaufmann H. Examination of young children with Lea symbols. *Br J Ophthalmol*. 2002;86:513-516.
- Friendly DS. Preschool visual acuity screening tests. *Trans Am Ophthalmol Soc*. 1978;76:383-480.
- Harvey EM, Dobson V, Tung B, Quinn GE, Hardy RJ. Interobserver agreement for grating acuity and letter acuity assessment in 1 to 5.5-year-olds with severe retinopathy of prematurity. *Invest Ophthalmol Vis Sci*. 1999;40:1565-1576.
- Hered RW, Murphy S, Clancy M. Comparison of the HOTV and Lea Symbols charts for preschool vision screening. *J Pediatr Ophthalmol Strabismus*. 1997;34:24-28.
- Kastenbaum SM, Kepford KL, Holmstrom ET. Comparison of the STYCAR and lighthouse acuity tests. *Am J Optom Physiol Opt*. 1977;54:458-463.
- Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci*. 1991;32:422-432.
- Raasch TW, Bailey IL, Bullimore MA. Repeatability of visual acuity measurement. *Optom Vis Sci*. 1998;75:342-348.
- Vanden Bosch ME, Wall M. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye*. 1997;11:411-417.
- Carkeet ABP. Modeling logMAR visual acuity scores: effects of termination rules and alternative forced-choice options. *Optom Vis Sci*. 2001;78:529-538.
- Elliott DB, Sheridan M. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. *Ophthalmic Physiol Opt*. 1988;8:397-401.
- Carkeet AP, Lee LB, Kerr JRB, Keung MMB. The slope of the psychometric function for Bailey-Lovie letter charts: defocus effects and implications for modeling letter-by-letter scores. *Optom Vision Sci*. 2001;78:113-121.
- Rosser DA, Murdoch IE, Cousens SN. The effect of optical defocus on the test-retest variability of visual acuity measurements. *Invest Ophthalmol Vis Sci*. 2004;45:1076-1079.
- Beck RW, Moke PS, Turpin AH, et al. A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol*. 2003;135:194-205.
- Lovie-Kitchin JE, Brown B. Repeatability and intercorrelations of standard vision tests as a function of age. *Optom Vis Sci*. 2000;77:412-420.
- Blackhurst DW, Maguire MG. Reproducibility of refraction and visual acuity measurement under a standard protocol. The Macular Photocoagulation Study Group. *Retina*. 1989;9:163-169.
- Reeves BC, Wood JM, Hill AR. Reliability of high- and low-contrast letter charts. *Ophthalmic Physiol Opt*. 1993;13:17-26.

23. Khetarpal S, Jones HS, Auld R, Moseley MJ. Reliability of visual acuity in children with reduced vision. *Ophthalmic Physiol Opt.* 1996;16:447-449.
24. McGraw PV, Winn B, Gray LS, Elliott DB. Improving the reliability of visual acuity measures in young children. *Ophthalmic Physiol Opt.* 2000;20:173-184.
25. Manny RE, Hussein M, Gwiazda J, Marsh-Tootle W. Repeatability of ETDRS visual acuity in children. *Invest Ophthalmol Vis Sci.* 2003;44:3294-3300.
26. Stewart C. Comparison of Snellen and log-based acuity scores for school-aged children. *Br Orthopt J.* 2000;57:32-38.
27. Sprague JB, Stock LA, Connett J, Bromberg J. Study of chart designs and optotypes for preschool vision screening. *Am Orthopt J.* 1988;38:18-23.
28. Vision in Pre-schoolers Group (VIP). Preschool visual acuity screening with HOTV and Lea symbols: testability and between-test agreement. *Optom Vis Sci.* 2004;81:678-683.
29. Cyert L, Schmidt P, Maguire M, Moore B, Dobson V, Quinn G. Threshold visual acuity testing of preschool children using the crowded HOTV and Lea Symbols acuity tests. *J AAPOS.* 2003;7:396-399.
30. Hyvarinen L, Nasanen R, Laurinen P. New visual acuity test for pre-school children. *Acta Ophthalmol (Copenb).* 1980;58:507-511.
31. Camparini M, Cassinari P, Ferrigno L, Macaluso C. ETDRS-fast: implementing psychophysical adaptive methods to standardized visual acuity measurement with ETDRS charts. *Invest Ophthalmol Vis Sci.* 2001;42:1226-1231.
32. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135-160.
33. Ferris FL 3rd, Bailey I. Standardizing the measurement of visual acuity for clinical research studies: guidelines from the Eye Care Technology Forum. *Ophthalmology.* 1996;103:181-182.
34. Graf MH, Becker R, Kaufmann H. Lea symbols: visual acuity assessment and detection of amblyopia. *Graefes Arch Clin Exp Ophthalmol.* 2000;238:53-58.
35. Ferris FL 3rd, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol.* 1982;94:91-96.
36. Repka MX, Beck RW, Holmes JM, et al. A randomized trial of patching regimens for treatment of moderate amblyopia in children. *Arch Ophthalmol.* 2003;121:603-611.