1040-5488/98/7505-0342$03.00/0 VOL. 75, NO. 5, PP. 342–348
OPTOMETRY AND VISION SCIENCE

ORIGINAL ARTICLE

Repeatability of Visual Acuity Measurement

THOMAS W. RAASCH, OD, PhD, FAAO, IAN L. BAILEY, OD, MS, FAAO, and MARK A. BULLIMORE, MCOptom, PhD, FAAO

The Ohio State University College of Optometry, Columbus, Ohio (TWR, MAB), and University of California School of Optometry, Berkeley, California (ILB)

ABSTRACT: *Purpose.* This study investigates features of visual acuity chart design and acuity testing scoring methods which affect the validity and repeatability of visual acuity measurements. *Methods.* Visual acuity was measured using the Sloan and British Standard letter series, and Landolt rings. Identifiability of the different letters as a function of size was estimated, and expressed in the form of frequency-of-seeing curves. These functions were then used to simulate acuity measurements with a variety of chart designs and scoring criteria. *Results.* Systematic relationships exist between chart design parameters and acuity score, and acuity score repeatability. In particular, an important feature of a chart, that largely determines the repeatability of visual acuity measurement, is the amount of size change attributed to each letter. The methods used to score visual acuity performance also affect repeatability. *Conclusions.* It is possible to evaluate acuity score validity and repeatability using the statistical principles discussed here. (Optom Vis Sci 1998; 75:342–348)

Key Words: visual acuity, visual acuity charts, optotypes, scoring methods, measurement validity, measurement repeatability

The basic principles of visual acuity measurement have remained essentially unchanged since Snellen introduced them in 1862[1]. Acuity charts typically consist of upper case letters within rows of decreasing size from top to bottom. Charts of this type are very commonly used to measure acuity, primarily because the test is readily administered with relatively simple equipment, and the results can provide valuable insights into the state of the eye's optics and the visual pathways. Clinical acuity measurement is probably the most commonly measured visual psychophysical threshold. Acuity is often considered to be the single most important indicator of the quality of vision. For example, letter acuity has been used as a primary outcome variable in large clinical trials, as in the Early Treatment Diabetic Retinopathy Study (ETDRS),[2] and in the Macular Photocoagulation Study.[3] Acuity is often the only visual function used to predict visually guided performance, such as in automobile driver licensing. It is likely that acuity performance will continue to be relied upon for a wide variety of applications. The pervasive use of acuity testing emphasizes the importance of understanding the effects that particular acuity chart designs and scoring methods have on acuity scores.

In 1980, the Committee on Vision, Working Group 39 of the National Research Council (NAS-NRC), issued a report that addressed several aspects of clinical visual acuity measurement.[4] The report stated the need for standards in visual acuity measurement, and made several recommendations. One goal of this working group was to improve the reliability of acuity testing, enabling examiners to obtain repeatable results that could be reasonably compared with the results of others. They identified nine issues that are amenable to standardization, and four of these issues are addressed in this paper. These four are: (1) comparison of optotypes to the Landolt ring; (2) the gradation and range in optotype size; (3) the number of symbols at each size level; and (4) the method of scoring.

The 1980 Committee on Vision did not raise the issue of *repeatability* of acuity scores. Because repeatability is a critically important characteristic of any clinical test, we focus primarily on repeatability when considering the effects of different chart designs and scoring methods.

Early work in the area of visual acuity chart design was primarily concerned with the design and selection of optotypes to be used in acuity testing (for a review, see Bennett[1]). One of the principal goals was to select a family of optotypes of nearly equal legibility. Sloan et al.[5] evaluated three test target types. She recommended a set of letters that could serve as a suitable substitute for the Landolt ring. The recommended letter set (the so-called "Sloan" letters) is composed of 10 nonserifed, uppercase letters formed within a square outline, with a stroke width of one-fifth the letter height.

Thomas W. Raasch is an AOF Ezell Fellow.

The British Standards Institution recommended a different letter series. The selection of these letters was based on the work of Coates[6] and of Woodruff.[7] The British Standard letters consist of a set of 10 nonserifed, uppercase letters, with a height/width ratio of 5/4, and a stroke width that is one-fifth the letter height. The members of these two letter sets are:

Sloan letters:	C D H K N O R S V Z
British Standard Letters:	D E F H N P R U V Z

It has long been assumed that nearly equal legibility of letters is necessary in order to produce reliable acuity scores.[1, 5, 6–8] We evaluate whether commonly used letter series are sufficiently similar in legibility so as not to inflate variability of acuity measures.

Most authors agree that a logarithmic size progression offers the greatest practical and theoretical advantages.[9–12] Sloan[10] proposed that resolution thresholds are similar to other visual thresholds in that they probably follow a Weber's law-like function. If resolution thresholds do indeed obey Weber's law, then equal increments in acuity would be represented by equal logarithmic steps in letter size. In other words, letters that are just noticeably different in size will be different by a constant proportion, regardless of the letter size.[12]

Most existing charts which do use a logarithmic size progression use a value of 0.1 log units (a ratio of 1.26×) between adjacent lines on the chart.[9, 11, 13] Letters on such charts approximately double in size every 3 lines, and increase 10-fold in 10 lines. An example of the commonly used size progression of 0.1 log units per size level is:

. . . 20 25 32 40 50 63 80 100 125 160 200 . . .

This particular choice of size progression has proven to be convenient, but is nonetheless arbitrary. Therefore, we evaluate the influence that size progression has on the chart score and score repeatability.

A chart design feature that is closely related to size progression is the number of letters per size level. The size progression and number of letters per line determine the total number of letters that must appear within a chart in order to span a given size range. We demonstrate that these two features, i.e., the size progression and the number of letters at each size level, are the features of a chart that are primarily responsible for determining the variability of acuity measurement with that chart.

The method chosen to score acuity performance also has an influence on the score, as well as the repeatability of that score. Scoring methods that assign scores in "whole-line" increments sacrifice repeatability.[11, 14–16] We show that the beneficial features of this type of scoring protocol extend across a wide range of chart design parameters.

Finally, several prior reports state there is an upper bound to the repeatability of acuity measurement because of the limitations of commonly available acuity charts.[15] We show, however, that it is possible to improve repeatability, if it is possible to simply increase the number of trials in a measurement of acuity.

MATERIALS AND METHODS
Empirical measurement of visual acuity

High contrast, black-on-white visual acuity charts were photographically prepared. Landolt rings, the 10 British Standard letters, and the 10 Sloan letters were used. For each optotype, three pairs of essentially equivalent charts were produced. The series consisted of nine chart pairs (3 letter sets by 3 spacings). The spacing between successive rows was equal to the between-letter spacing of the larger row. The three spacings were 0.8, 1.0, and 1.25× the letter width. Each row contained five letters, and there were seven rows of letters in each chart. The size progression between successive rows was 0.1 log units. For charts in an "equivalent" pair, only the letter sequencing was different between the two charts. These chart specifications are similar to the well-known Bailey-Lovie and ETDRS visual acuity charts.[9, 11, 17] At the test distance of 4 m, the letter heights spanned visual angles from 10 to 2.5 min arc (6/12 to 6/3 equivalent, or 20/40 to 20/10 equivalent). All testing was done under incandescent illumination, and the photopic luminance of the white background of the letters was 160 cd/m^2.

Subjects

Nineteen normally sighted subjects were tested. The subjects were volunteers from the student population of the School of Optometry at the University of California, Berkeley. The age range was from 20 to 33 years. All subjects had visual acuity of 20/25 (6/7.5) or better, were free of any known eye disease, and each used his or her habitual correction. Rigid contact lens wearers were excluded, and soft contact lenses wearers were allowed to participate if at least 12 h had elapsed since they had last worn their lenses.

Visual acuity scoring procedure

The subjects were asked to read each chart, starting with the largest row. Subjects were permitted to stop reading a chart when they missed three or more of the five letters in a row. The subjects' responses were recorded on facsimiles of the charts, noting which letters were read correctly. The experimental sessions lasted approximately 20 min. Charts were presented in a random sequence.

A visual acuity score was obtained for each chart by assigning a value of logMAR = 0.02 to each letter. The top row subtended a visual angle of 10 min arc, or MAR of 10/5 = 2 min arc, so if the subject read this entire line and no more, he/she would be assigned an acuity of 20/40, or 0.30 logMAR. Each additional letter read would improve his/her acuity score by 0.02 log units. The rationale for this procedure has been described elsewhere.[11, 14, 18]

Modeling of acuity measurement with simulations

The empirical data described above were used to model and simulate acuity performance. Probability of seeing curves for each letter in the two letter series were computed. The threshold size for each subject on each chart was determined using the single-letter scoring criterion described above. To construct the probability of seeing curves, letter size was adjusted according to the threshold size obtained by that subject on that chart. Data were accumulated across subjects to build probability of seeing values for each of the

10 letters in the Sloan and British Standard letter series. These probability of seeing data were then fit with cumulative normal distribution curves using Probit analysis.[19] This process is similar to that used by others[8, 15, 16] in describing acuity performance.

These curves represent the probability that a letter will be identified correctly as a function of its size relative to the threshold that that subject obtained on that chart. Using these probability of seeing curves, it is possible to simulate acuity performance across a wide range of chart parameters. The effects on acuity scores and repeatability are studied by manipulating the size ratio between lines, and the number of letters per line, and then simulating acuity performance, using the probability of seeing functions. Charts used in the simulations were constructed so that no letters appeared more than once at a given size level, all charts contained seven rows of letters, and the size of the chart was randomly set so that the height of the letters in the middle line subtended an angle somewhere between 4 and 6.3 min arc. For each simulated chart condition, 1000 versions were generated and scored.

Derivation of scores and score variances from simulated acuity measurement

The probability that a letter in a chart, c_i, will be identified correctly has a value between 0.0 and 1.0, and the expected score for c_i is represented by $p(c_i)$. Likewise, the probability that this letter will *not* be read correctly is $q(c_i) = 1 - p(c_i)$. The expected score for the entire chart, in the number of correctly read letters, is the sum of these individual letter probabilities, $\Sigma_i p(c_i)$.

Furthermore, the variance of the score, in letters^2 can be estimated. For a single letter, the variance of the expected score for that letter is $\sigma_i^2 = p(c_i) \times q(c_i)$. This single letter variance ranges between 0.0 letters^2 (for letters much above and much below threshold size) to 0.25 letters,^{2} for letters with $p(c_i) = 0.5$. The variance of the score for the whole chart is the sum of the variances of the letters making up the chart, $V = \Sigma_i \sigma_i^2$. The SD of the chart score, S, is the square root of this variance, $\sqrt{V}$. Furthermore, in a test-retest experiment in which the test and retest measurements are independent, the SD of the discrepancy score is $D = \sqrt{2} \times S$.

Fig. 1 is an example of a chart consisting of Sloan letters, in which the letters span a four-fold size range. Included under each letter is the probability of correctly identifying the letter and the variance of that probability. Given these particular letters and sizes, this particular chart has an expected score of 19.975 letters, a variance of that score of 1.67 letters,^{2} and a SD of the score of 1.292 letters. In this chart each letter accounts for 0.02 log units, so this score SD is: 1.292 letters × 0.02 log units/letter = 0.02584 log units. Of course, if this chart is viewed by a different observer, perhaps with a different acuity, and from a different distance, these probabilities and variance will differ.

						Angular Size
	C	K	D	N	R	
Probability:	1.000	1.000	1.000	1.000	1.000	10.0
Variance:	0.000	0.000	0.000	0.000	0.000	min arc
	S	R	Z	K	D	
Probability:	1.000	1.000	1.000	1.000	1.000	8.0'
Variance:	0.000	0.000	0.000	0.000	0.000	
	H	Z	O	V	C	
Probability:	1.000	1.000	0.972	1.000	0.993	6.3'
Variance:	0.000	0.000	0.027	0.000	0.007	
	N	V	D	O	K	
Probability:	0.780	0.924	0.817	0.492	0.739	5.0'
Variance:	0.171	0.070	0.149	0.250	0.193	
	V	H	C	N	O	
Probability:	0.306	0.572	0.076	0.122	0.025	4.0'
Variance:	0.212	0.245	0.070	0.107	0.024	
	S	V	H	C	Z	
Probability:	0.000	0.007	0.039	0.000	0.111	3.2'
Variance:	0.000	0.007	0.038	0.000	0.099	
	O	Z	D	V	K	
Probability:	0.000	0.000	0.000	0.000	0.000	2.5'
Variance:	0.000	0.000	0.000	0.000	0.000	

FIGURE 1.

Example of a chart composed of Sloan letters. Angular size is listed to the right, and the acuity of the observer is arbitrarily set to be approximately 20/20. The probability and variance values are listed under each letter. For this particular chart, the expected score is the sum of the individual letter probabilities, which is 19.975 letters (0.0005 logMAR), and the variance of this score is the sum of the individual letter variances, which is 1.669 letters.^{2} The SD is therefore 1.292 letters, or 0.026 log units. Note that most of the variance in the score is contributed by near-threshold letters.

TABLE 1.

Acuity and discrepancy in the three families of optotypes, averaged across subjects.

Letter Family	Acuity Relative to Landolt Ring			Test-Retest Discrepancy (SD)	
	LogMAR	Letters	Lines	Letters	Log units
Landolt Rings	—	—	—	2.5	0.050
British Letters	−0.005	−0.25	−0.05	1.8	0.036
Sloan Letters	−0.038	−1.9	−0.38	2.4	0.047

RESULTS

Comparison of empirical acuities with letter optotypes and Landolt rings

These data indicate there are small differences among acuities obtained with the letter optotypes and Landolt rings (see Table 1). Similar results are reported elsewhere.[14] Acuities obtained with Sloan letters are slightly better than those obtained with Landolt rings (better by 0.038 log units, or almost 2 "letters" better with Sloan letters). Sloan[5] obtained data that suggested to her that the difference between acuities obtained with Sloan letters and Landolt rings is not clinically significant. The difference between acuities obtained with the British Standard letters and with Landolt rings is very small, i.e., 0.005 log units (or 0.25 letters) better acuity with the British letters.

Included in Table 1 is an indication of the width of the distribution of test-retest discrepancy scores. The discrepancy scores for the Sloan letters and the Landolt rings are very similar, and both are larger than for the British Standard letters. These data suggest that charts comprised of British Standard letters will produce slightly more repeatable acuity measurements than charts comprised of either Landolt rings or Sloan letters.

Table 2 lists the acuity and test-retest discrepancy values, averaged across subjects for the three spacing arrangements. The aver-

TABLE 2.
Visual acuity (relative to spacing = 1.0) and test-retest discrepancy scores for the three spacing arrangements, averaged across subjects and British and Sloan letter sets.

Spacing (× letter width)	Acuity		Discrepancy SD	
	LogMAR	Snellen notation	Letters	Log units
0.8×	0.009	20/20.4 (6/6.12)	2.40	0.048
1.0×	—	20/20 (6/6)	2.05	0.041
1.25×	−0.014	20/19.4 (6/5.8)	2.10	0.042

age acuity level improves with increasing spacing between letters, and this was expected from the results of other studies. Discrepancy values, however, do not change consistently with spacing. That is, although acuity level does change systematically with spacing, repeatability does not.

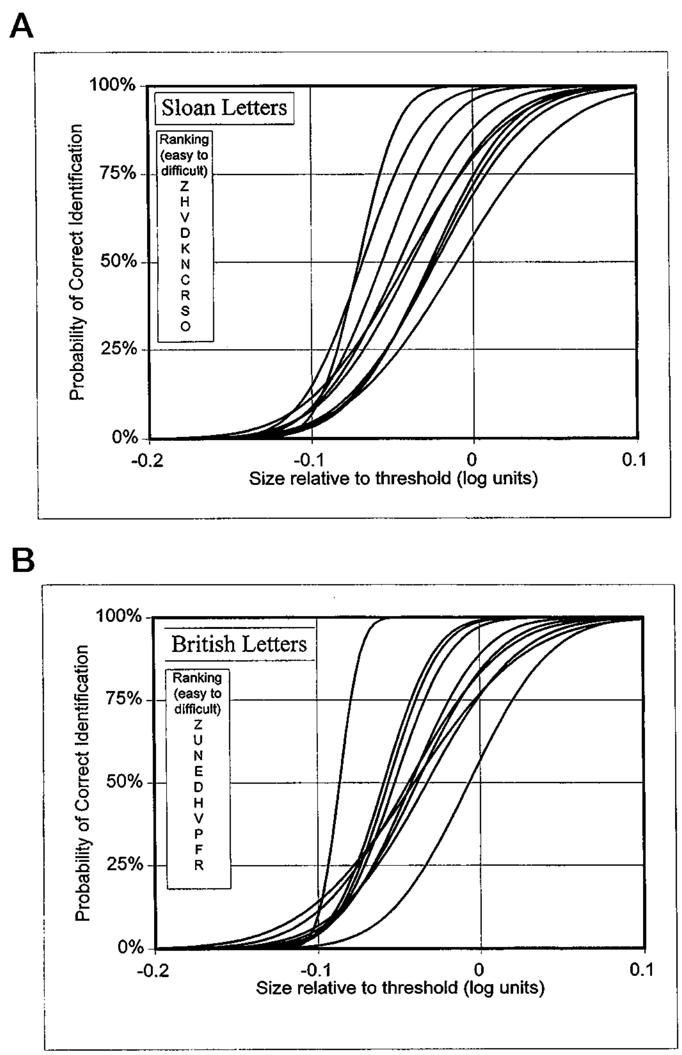

FIGURE 2.
Probability of correct identification as a function of letter size. Letter size is expressed relative to threshold. A: Sloan letter series. B: British Standard letter series.

Probability of correct identification of letters

The experimental data were used to construct probability of correct identification curves, as shown in Fig. 2, A and B. From these curves, acuity performance was simulated under a variety of chart design formats. The 10 curves in each figure specify the probability that a letter will be identified correctly, given its size relative to threshold. Their horizontal positions reflect their different threshold sizes for a given percent correct. Their vertical positions reflect their difficulty (or percent correct), at a given size. The difficulty rankings listed in this figure are based on percent correct levels at threshold size. Some curves cross others, so those rankings may shift if a different size level is selected. This issue is discussed by Alexander et al.[8] As pointed out by Alexander et al., the spectral power of the object spatial frequencies of these letters differs, and so their relative identifiability can change, depending upon their angular size.

Effects of size progression and letters per line

Fig. 3 illustrates the results of a simulation of acuity performance, in which size progression and letters per line were systematically varied. Fig. 3 shows that acuity score does depend upon size progression between adjacent lines on the chart, but does not depend upon the number of letters at each size level, i.e., the 10 lines are superimposed. These lines have a slope near ½, indicating that as size progression increases by n log units, the acuity score increases (i.e., gets worse) by n/2 log units.

Fig. 4 shows that the repeatability of the chart score does depend upon both the size ratio and the number of letters at each size. These plots on log-log axes have slopes near ½, suggesting a square root relationship. That is, as the size progression between lines increases by a factor of n, the SD of the score increases by a factor of $\sqrt{n}$.

Fig. 5 is derived from Fig. 4. Fig. 5, however, combines size ratio

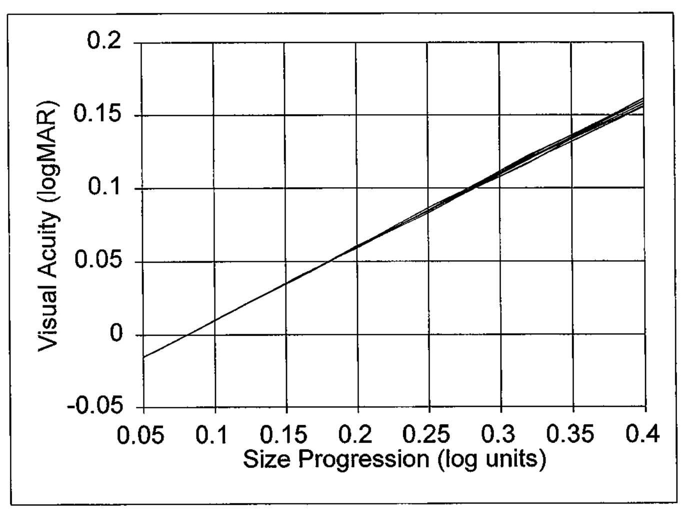

FIGURE 3.
Acuity thresholds as a function of size progression and number of letters per line. Sloan letters. There are 10 lines, derived from simulations of 1, 2, 3. . . 10 letters per line. The 10 lines are superimposed, indicating that the chart score is independent of the number of letters per line. The slope of 0.5 indicates that as size progression changes by an increment of n units, chart score will change by $\sqrt{n}$ units.

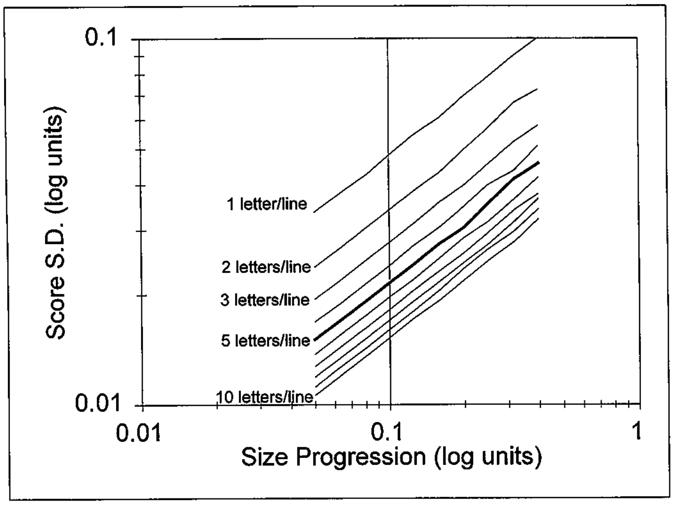

FIGURE 4.
SD of the acuity score as a function of size progression and number of letters per line. Sloan letters. The bold line represents the commonly used five letters per line. The slope of 0.5 indicates that as size progression changes by a factor of n units, chart score will change by a factor of $\sqrt{n}$ units.

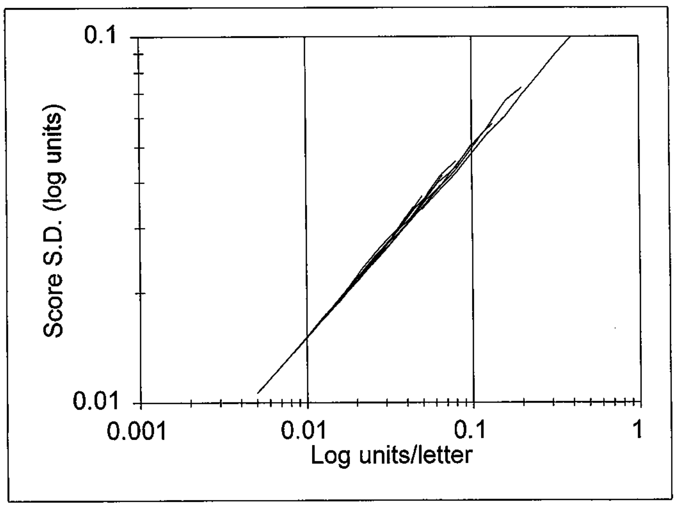

FIGURE 5.
SD of the acuity score expressed as a function of the number of log units accounted for by each letter on the chart. Sloan letters. These 10 lines are superimposed, indicating that regardless of the number of letters per line, the repeatability of the score is determined primarily by how many log units of size change are attributed to each letter.

and letters per line into a single variable. The size progression per line, divided by the number of letters per line, yields log units of size progression accounted for by each letter. For example, the common chart format of 0.1 log units per line, and five letters per line, yields:

$$\frac{0.1 \text{ log units/line}}{5 \text{ letters/line}} = 0.02 \text{ log units/letter.}$$

Fig. 5 illustrates an important concept: the repeatability of an acuity measurement is determined by the number of log units in size change accounted for by each letter. This holds true across a wide range of letters per line and size progressions.

Scoring criteria

Scoring criteria have been evaluated by a number of authors in recent years.[11, 14–16, 20] The issue is whether "whole line" or "individual letter" scoring methods yield more repeatable acuity scores. All agree that a scoring method that awards credit in increments of single letters is better than a method that awards credit in whole line increments. The present findings are in agreement with these other studies. The standard error of the visual acuity measurement using whole line scoring methods tends to yield values that are approximately 1.6 to 1.7 times larger than individual letter scoring methods. This difference between individual letter and whole line scoring methods stays at this level, across a wide range of size progressions and letters per line.

DISCUSSION

Reliability of acuity scores

If acuity is tested twice on essentially equivalent charts, the expected difference in score is zero. If tested many times, there will be a distribution of difference scores, and this distribution reflects the underlying variability in the acuity measurement. These types of test-retest data have been analyzed in at least two ways, including use of the Intraclass Correlation Coefficient,[21] and test-retest discrepancy values, or the coefficient of reliability, as described by Bland and Altman.[22] We believe that test-retest discrepancy scores are an appropriate type of analysis, as it is a direct indication of the within-subject variability. The Intraclass Correlation Coefficient does incorporate a term for within-subject variance, but it is also affected by the range of the acuities tested, i.e., the between-subject variance. If results are to be compared across experiments, the Intraclass Correlation Coefficient may obscure the comparison, because between-subject variance may mask differences in within-subject variance.

There is, of course, a relationship between within-subject variance and test-retest discrepancy scores. The Intraclass Correlation Coefficient, R, is equal to:

$$R = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2},$$

where σ_T^2 is the variance of the mean values of the n subjects, i.e., the between-subject variance, and σ_e^2 is the within-subject variance.[21] As the between-subject variance increases, R will also tend to increase, illustrating that this statistic is affected by the range of mean values. When the mean discrepancy score is zero (i.e., when there is no overall trend from test to retest), the variance of the discrepancy scores is: $\sigma_D^2 = 4(\sigma_e^2)$.

A number of other authors in recent years have studied repeatability of visual acuity measurement.[11, 14–16, 20] There are small differences in the results reported by these different authors, as the data sets are the result of different experimental methods and subject populations. However, the differences in repeatability found for visual acuity measurement are not great. Table 3 lists repeatability data from several studies in which Sloan letters were used, and with a size progression of 0.1 log units and five letters per size level. Most studies find that test-retest SDs are on the order of two to three letters (i.e., 0.04 to 0.06 log units).

TABLE 3.
Repeatability of visual acuity testing reported in the literature.[a]

	Bailey & Lovie (1976)[9]	Elliott & Sheridan (1988)[20]	Bailey et al. (1991)[14]	Arditi & Cagenello (1993)[15]	Vanden Bosch & Wall (1997)[16]	Raasch et al. (present study)
SD of VA Score	1.8	1.93	1.4	2.16	1.9	1.66
Test-Retest SD	2.55	2.73	1.98	3.06	2.69	2.35
95% Confidence Interval	5	5.35	3.88	6	5.27	4.61

[a] All scores from use of Sloan letters, a size progression of 0.1 log units, and five letters per size level. Units are letters (0.02 log units/letter).

Important chart design principles

Previous authors[4, 9, 11, 12] have argued in favor of a logarithmic size progression in letter size. Likewise, it has been proposed that the task be essentially the same at each size level.[9] This requires, among other things, that the same number of letters be present at each size level. When these two features are present, each letter in a chart accounts for an equal increment in acuity.

The slope of the function relating SD of the acuity score to the size change per letter is equal to ½, suggesting a square root rule. Indeed, this would be expected if an acuity score is simply considered a sample of the subject's acuity. According to a square root rule, increasing the number of letters used to measure the acuity by a factor of n will reduce the standard error of the estimate by a factor of $\sqrt{n}$. As Arditi and Cagenello[15] point out, the repeatability of an acuity measurement could be improved if the size progression were 0.05 log units, rather than the common 0.10 log units. This is not the only way to improve repeatabilty, however. For example, the number of letters could be doubled by doubling the number of letters in each line, or by measuring the acuity twice (with different but otherwise equivalent charts), and averaging the results. Either of these improves the precision of the estimate of acuity by a factor of $\sqrt{2}$.

The results of both the empirical data and the simulations comparing the four different scoring criteria indicate that acuity measurements are most accurate when measured in the finest size increments available, i.e., in single letter increments. This is due essentially to the fact that letter-by-letter scoring will yield relatively few instances in which the disparity is as large as a whole line, a result that arises relatively frequently with a whole-line criterion.[14] The end result is that the distribution of test-retest discrepancies (and hence the score SD) is narrower when an individual letter scoring method is used.

Is equal legibility of letters an advantage?

It has been assumed that in order to achieve the most reliable acuity measurements possible, the letters within a series should be equally legible.[4, 5, 8, 23] It is unlikely that perfectly equal legibility can be achieved, even with targets such as Landolt rings, where meridional differences may favor some orientations over others.[24] Recognizing that the available letter sets are not equally recognizable, what are the consequences? Are the differences in legibility sufficiently great to contribute to variability in acuity measurement? To answer these questions, two additional simulations were performed. In the first, simulated acuity measurement was performed using Sloan letters arranged in the common chart format of 5 letters per line and 0.1 log unit size progression. The second simulation was performed in the same way, after substituting a hypothetical letter set in which all had the same legibility. That is, the probability of identification curves (as in Fig. 2) was identical for all letters in this set, and was the average of those for the Sloan letters.

The results of these simulated acuity measurements indicate that Sloan letter chart scores are just slightly more variable than scores for the "equal legibility" chart. The Sloan chart SD was equal to 1.072 letters vs. 1.026 letters for the equal legibility chart. The difference is less than 1/20th of a letter, and suggests that, for most purposes, the inequality of identifiability of these letters is not a significant factor. It should be noted, however, that if a chart is made up primarily of "easy" or "hard" letters, scores obtained with that chart would be affected. In designing a chart, it is probably sufficient to simply use all letters in the set with about equal frequency, and more or less randomly. It is also probably unnecessary to pursue an alternative set of letters or symbols that are perfectly equal in legibility, as the benefits of even perfect equality are quite modest.

Evaluation of existing charts

Existing charts can be evaluated with these principles in mind. Some charts may have a format in which the size progression is as low as 0.012 log units per letter at the smaller letter sizes. The same chart may have a size progression at the larger letter sizes of 0.3 log units per letter, as when single letters appear at the 400 and 200 foot-letter sizes. The score variability is exceedingly high at this level, which is noteworthy because the threshold for determination of legal blindness is in this acuity region (see Fig. 5). This is an important reason why an acuity measurement with substantial consequences, such as establishment of legal blindness, should perhaps not be done with an acuity chart that has this weakness.

Clinicians have observed that some low vision patients exhibit large acuity differences from one measurement to another. Greater variability in acuity performance by some low vision patients may play a part in this,[25] but some acuity charts may contribute additional variability. Charts used for the measurement of low visual acuity necessarily include large characters, so in order to limit the overall size of these charts, relatively few letters may be used. Consequently, such charts often use large size ratios from one line to the next (to reduce the number of lines), and may have few letters on each line. Although this reduces the size of such charts, the effect is to make the measured acuities more variable. In order to achieve greater repeatability, the size increment accounted for by each letter should be kept as small as is practical, and preferably constant, as letter size is changed.

NAS-NRC standards for visual acuity measurement

The NRC Committee on Vision has offered several recommendations for the measurement of visual acuity.[4] The Landolt ring has been proposed as the standard optotype, although the committee states that other optotypes which yield similar acuities are acceptable. The empirical data of the two series of letter optotypes indicate that acuities obtained with the British Standard letters are very close to those obtained with Landolt rings. Acuities obtained with Sloan letters may be slightly different, yielding threshold sizes that are approximately 8% smaller than those obtained with Landolt rings or British Standard letters.

The NRC committee has recommended that a logarithmic size progression of 0.1 log units be used, and that all 10 letters in the series be presented at each size level. Existing charts infrequently include so many symbols at each size. Size progression has two effects. First, change in size progression by an increment of n will change the score by an increment of $\sqrt{n}$ units (see Fig. 3). Second, and more important, multiplying the size progression by a factor of n will multiply the width of the distribution of discrepancy scores by $\sqrt{n}$ (see Fig. 4 or Fig. 5).

The NRC committee has further recommended that a scoring criterion of 7 of 10 (or 6 of 8) letters be required to pass a size level. The results of this study and of others[11, 14, 15, 18, 20] indicate that these whole-line scoring methods do a relatively poor job in terms of score repeatability. A scoring method that awards credit in single letter increments produces scores that are more repeatable. Charts in which each letter accounts for a constant proportion of letter size are conducive to scoring methods that award credit in individual letter increments, and the use of such scoring methods is recommended for situations in which accurate and reliable acuity measurement is desired (see Fig. 5).

Finally, it has been suggested that there is an upper bound on the precision with which acuity can be measured. This is true if there are limitations placed on the measurement procedure, such as a shortage of time or the number of acuity charts or letter sequences available for use. If it is possible to measure acuity more than once, with charts that are equivalent (e.g., the same design except for letter sequencing), then the repeatability of the results can be enhanced by simply averaging the results. The average is better, in terms of the standard error of measurement, by a factor of $\sqrt{2}$. This is a consequence of the ½ slope in Figs. 4 and 5. Testing twice essentially halves the number of log units accounted for by each letter. If tested with still more letters, it is possible to enhance the repeatability of the measurement further.

ACKNOWLEDGMENT

Dr. Raasch was supported by an Ezell Fellowship during the period this study was conducted.

Received August 5, 1993; revision received January 5, 1998.

REFERENCES

1. Bennett AG. Ophthalmic test types. Br J Physiol Opt 1965;22: 238–71.
2. Early Treatment Diabetic Retinopathy Study Group. Photocoagulation for diabetic macular edema. Arch Ophthalmol 1985;103: 1796–806.
3. Macular Photocoagulation Study Group. Visual outcome after laser photocoagulation for subfoveal choroidal neovascularization secondary to age-related macular degeneration. The influence of initial lesion size and initial visual acuity. Arch Ophthalmol 1994;112:480–8.
4. National Research Council Committee on Vision. Recommended standard procedures for the clinical measurement and specification of visual acuity. Adv Ophthalmol 1980;41:103–48.
5. Sloan LL, Rowland WM, Altman A. Comparison of three types of test target for the measurement of visual acuity. Q Rev Ophthalmol 1952;8:4–16.
6. Coates WR. Visual acuity and test letters. Transactions of the Institute of Ophthalmic Opticians III, 1935.
7. Woodruff EW. Visual acuity and the selection of test letters. In: Some Recent Advances in Ophthalmic Optics: Transactions. London: Hatton Press, 1948:59–70.
8. Alexander KR, Wei X, Derlacki DJ. Visual acuity and contrast sensitivity for individual Sloan letters. Vision Res 1997;37:813–7.
9. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. Am J Optom Physiol Opt 1976;53:740–5.
10. Sloan L. The measurement of visual acuity. A critical review. Arch Ophthalmol 1951;45:704–25.
11. Ferris III FL, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. Am J Ophthalmol 1982;94:91–6.
12. Westheimer G. Scaling of visual acuity measurements. Arch Ophthalmol 1979;97:327–30.
13. Regan D. Low-contrast letter charts and sinewave grating tests in ophthalmological and neurological disorders. Clin Vision Sci. 1988; 2:235–50.
14. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. Invest Ophthalmol Vis Sci 1991;32: 422–32.
15. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. Invest Ophthalmol Vis Sci 1993;34: 120–9.
16. 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;11:411–7.
17. Strong G, Woo GC. A distance visual acuity chart incorporating some new design features. Arch Ophthalmol 1985;103:44–6.
18. Elliott DB, Bullimore MA, Bailey IL. Improving the reliability of the Pelli-Robson contrast sensitivity chart. Clin Vision Sci 1991;6:471–5.
19. Finney DJ. Probit Analysis. 3rd ed. Cambridge University Press, 1971.
20. Elliott DB, Sheridan M. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. Ophthalmic Physiol Opt 1988;8:397–401.
21. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: John Wiley & Sons, 1986.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307–10.
23. Hedin A, Olsson K. Letter legibility and the construction of a new visual acuity chart. Ophthalmologica 1984;189:147–56.
24. Mitchell DE, Freeman RD, Millodot M, Haegerstrom G. Meridional amblyopia: evidence for modification of the human visual system by early visual experience. Vision Res 1973;13:535–58.
25. Raasch TW, Flom RE. Precision of visual acuity measurement in low vision. Invest Ophthalmol Vis Sci 1994;35(Suppl):1413.

Thomas W. Raasch
The Ohio State University College of Optometry
338 W. 10th Avenue
Columbus, Ohio 43210