Improving the Design of the Letter Contrast Sensitivity Test

Aries Arditi

PURPOSE. To develop a small-format letter contrast sensitivity test with improved accuracy, suitable for portable use and near testing, and having a simple and unambiguous scoring system that can be used with reference to existing norms.

METHODS. A near test, printed on resin-coated paper and mounted on plastic was developed by using Sloan letters, with 0.04 log unit contrast decrements between each letter and a simple scoring procedure. Monte Carlo methods and a Weibull function model of visual performance were used to assess test accuracy.

RESULTS. The new test has 28% lower score standard deviations than the Pelli-Robson Contrast Sensitivity Chart, over a wide range of low misreporting rates, while maintaining test scores that differ from the Pelli-Robson by 0.01 log unit, permitting use of norms collected with the Pelli-Robson test.

CONCLUSIONS. While maintaining comparability with Pelli-Robson norms, the new test has improved accuracy in comparison with the Pelli-Robson chart and several other advantages that result from its smaller size. (Invest Ophthalmol Vis Sci. 2005; 46:2225–2229) DOI:10.1167/iovs.04-1198

Contrast sensitivity (CS), while itself associated with visual acuity, characterizes aspects of visual function that are not as well captured by clinical visual acuity measurement. It is especially strongly associated with reading performance,2,3 ambulation mobility,14–17 driving,8,9 face recognition,10,11 and tasks of daily living.11–13 It thus may be especially useful in determining disability for people with vision impairments. Recently, the National Research Council published a report that recommended adding CS as a basis for disability determination for the U.S. Social Security program.14 There is mounting evidence in the literature that suggests that CS may also be a sensitive indicator of disease and disease progression.15–29

CS has traditionally been measured with gratings.30,31 With this method, sensitivity is measured to a sample of spatial frequencies that span the visible range. But in recent years, letter CS, which measures CS with large letters and only in the spatial frequency band where CS is peak, has been the preferred method of CS measurement in clinical settings. There are several reasons for this: First, peak CS and ordinary visual acuity alone are sufficient to determine the full form of the CS function in all but a small number of controversial cases.32,33 Given this and the fact that visual acuity measurement is routine, to obtain a full CS curve would be a potential waste of clinical time. Second, letter CS is easy for the clinician or technician to perform, and patients quickly and easily grasp the letter-reading task, which is similar to that used in measuring Snellen acuity. Third, it has good repeatability,34–37 making it suitable for use in large populations,38–40 as well as in smaller clinical studies.17,21–24,41–43

Until now, the only available test of this kind has been the Pelli-Robson Contrast Sensitivity Chart.44 This report describes a new test, the Mars Letter Contrast Sensitivity Test, whose design is similar to but improves on the Pelli-Robson chart. Because the Pelli-Robson chart is widely known, a brief description will be given of that chart’s characteristics, along with a discussion of several variants of scoring procedures that have been used (but often not documented), followed by a description of the improvements of the Mars test.

The Pelli-Robson chart, originally described by Pelli et al.,44 is a large wall-mounted chart, 59 cm wide and 84 cm high, that consists of 16 tripels of Sloan letters5 each subtending 2.8° at the recommended 1-m test distance, arranged in eight rows of two tripels each. The three letters within each triplet have constant contrast, whereas the contrast across tripels, reading from left to right and continuing on successive lines, decreases by a constant factor (1/√2 or 0.15 log unit). The patient reads the letters across and down the chart, as in standard letter acuity measurement. Instead of the letters decreasing in size, however, they decrease in contrast. The final triplet at which the patient reads at least two of three letters correctly determines the log CS, which can be obtained from a score sheet that relates each triplet to a log CS value. This is the manufacturer-recommended scoring method.

The Pelli-Robson Chart available now from Haag-Streit (Köniz, Switzerland) departs somewhat from the one described by Pelli et al.44 and from the one first made available in the early 1990s. At that time, the charts were printed on two sides of a single sheet of plastic, requiring that it be flipped over to use both forms, whereas the current one is printed on two separate sheets of resin-coated paper mounted on card-board, which together require almost 1 m² of wall space. Test distance in the original Pelli et al.44 paper was specified as 3 m, with the letters subtending 0.5°, whereas with early commercial versions instructions were included for testing at 1 m (with somewhat larger letters, now subtending 2.8°). Recommended test distance is now 1 m, with 2.8° letters. Log CS originally ranged from 0.05 to 2.30, whereas in the currently available chart, it now ranges from 0.00 to 2.25. Finally, because the contrast values may change over time due to exposure to light, the chart is now sold with an expiration date beyond which the chart’s accuracy is uncertain.

Changes in chart design over the years may make interpretation of norms difficult, but even more challenging to interpret are the many combinations of scoring rules and stopping procedures that users have adopted. Elliott et al.,46 early on, using Monte Carlo methods similar to those described in the present paper and similar to the analytic derivation of accuracy...
originally used by Pelli et al.,44 showed that scoring the test by giving 0.05-log unit credit for each letter improves the test’s reliability and reduces the standard deviation of test scores; and, based on these results, this practice has been used in many studies. Elliott et al. adopted the stopping rule of terminating the test when none of the letters of a triplet was correctly identified. Although some studies, such as the Smith-Kettlewell Institute (SKI) longitudinal study 40 have adopted both the scoring and stopping rule of Elliott et al.,46 many use different stopping rules and scoring changes, often without reporting which stopping rule was used at all. The Salisbury Eye Evaluation (SEE) study,38 for example, used letter-by-letter scoring, but terminated the test when two letters of a triplet were incorrectly identified (Rubin GS, personal communication, July 2, 2004), which is the same stopping (but not scoring) rule recommended by the manufacturer and by Pelli et al.,44 but not by Elliott et al.,46 the originators of the letter-by-letter scoring method. Haymes et al.47 used the same two-of-three-errors stopping rule,46,48 but after suggestions by Elliott et al.,46 who recommended accepting the response “O” for a presented “C” (but not vice versa), accepted either “O” or “C” for presentation of either letter (Haymes S, personal communication, July 2, 2004). Thus, although some investigators have followed the manufacturer’s instructions,49,50 there are many variants of stopping and scoring rules in common use, often not specified in publications and reports. These complicate interpretation, as will be shown herein, for two reasons. First, each stopping–scoring variant produces a different test accuracy (see Fig. 2). Second, each variant converges on a different CS value (see Fig. 3).

The Pelli-Robson chart has been widely used by researchers for many years now, but it has not been widely adopted by clinicians. There are several probable, very practical reasons for this. In addition to clinicians’ understandable aversion to using an instrument that has no agreed-on or consistent rules for use and the fact that visual acuity is the required standard for Medicare and most third-party reimbursemens. First, it is inconvenient for testing in small clinical spaces, as it requires the devotion of a large amount of wall space. Second, it is difficult to arrange lighting that will illuminate such a large area uniformly (especially in a small room, since it is difficult to provide even light with near light sources). Third, a chart mounted at eye level (when the patient is seated), is difficult to keep clean and free of defects in a busy and heavily traveled examination room. Finally, unless the examination room wall happens to be 1 m from the examination chair, the patient has to move to the testing location, which, especially with elderly patients, can be a time-consuming and inconvenient affair.

**METHODS**

The Mars Letter Contrast Sensitivity Test, shown in Figure 1, is printed on resin-coated paper by halftone screening methods similar to those used in the Pelli-Robson Contrast Sensitivity Chart. It measures 9 × 14 in. (22.8 × 35.6 cm), and is mounted on plastic. Three forms of the chart are supplied (for left eye, right eye, and binocular testing, or for repeated testing), and the user manual is available in English, Spanish, and Chinese (Mandarin).

The test is intended for testing at 0.5 m, at which distance each letter subtends 2°. It can be used at a more conventional 40 cm near refraction distance as well, where letters subtend 2.5°, nearly the same as those on the Pelli-Robson. It can be handled by either the patient or the examiner. Alternatively, it can be placed on a stand on a table, or mounted with Velcro. It does not require significant office space. When not in use, it can be stored in its plastic case, which protects it from the harmful effects of light and potential mechanical damage.

Because of its small size, it is easy to illuminate the chart uniformly, typically with an instrument stand lamp angled at 45°. The recommended illumination is 85 cd/m², the same as recommended by the Committee on Vision of the National Research Council 51 and the same as the Pelli-Robson chart. In conjunction with portable lighting, the test can be used for vision function testing within patients’ homes or other atypical locations.

The Mars Letter Contrast Sensitivity Test follows many of the same design principles as the Pelli-Robson test. 43 It uses the same Sloan letter forms and presents the letters in declining contrast across and down the chart. The Mars test, however, uses much smaller contrast decrements (0.04 log unit) than the Pelli-Robson (0.15 log unit); these are applied letter by letter (thus there are no triplets). The test stops when the patient makes two consecutive errors. The score is the log CS of the final correct letter, minus 0.04 for any errors before that. This scoring procedure is simple and consistent and has the advantages of letter-by-letter scoring built in.

I used analytic and Monte Carlo methods and the very same simple model of visual performance used by Pelli et al.,44 to compare performance on the Mars test with that of the Pelli-Robson, including several variants of its scoring methods. This model assumes that the observer’s threshold is modeled by a Weibull psychometric function. Specifically, the probability of correctly naming a letter at log contrast, $x$ is given by

$$P(x - t) = (1 - \lambda)(1 - (1 - g) \exp[-10^{\beta(x - t)}] + g\lambda,$$

where $t$ is the observer’s threshold log Weber contrast (assumed to be −2.0 for the simulations, the same value Pelli et al.44 used), $g$ is the guessing probability (fixed at 0.1 for the case of the 10 Sloan letters), $\beta$ is a parameter that governs steepness of the function (which is fixed at 3.5 for the simulations—again the same value Pelli et al.44 used), and $\lambda$ is the lapse or misreporting rate, which, as noted by Pelli et al.,44 is the only parameter that has a significant impact on test score or test score variability. The Weber contrast is defined as $(I_{bg} - I_{lg})/I_{bg}$ where $I_{bg}$ and $I_{lg}$ are background and foreground luminance, respectively. The misreporting rate is the probability that the observer gives an incorrect response for reasons having nothing to do with sensory limitations. For example, an observer might simply be inattentive, or make a letter pronunciation error. Misreporting has nothing to do with threshold, except insofar as it can add noise to, or bias estimates of, threshold. In each simulated chart reading, responses to each letter were generated by sampling from a binomial distribution with probability $P(x - t)$ computed from the equation just shown, for the contrast associated with each letter. These simulated chart readings were then scored by either the Mars test method or one of the Pelli-Robson methods.
RESULTS

Figure 2 compares the accuracy of the Mars test, as characterized by test score standard deviation, with that of the Pelli-Robson scored three different ways: (1) with manufacturer-recommended scoring (w/MRS); (2) with letter-by-letter scoring with a stopping rule of two incorrect within a triplet (LBL 2/3); and (3) with letter-by-letter scoring with a stopping rule of three incorrect within a triplet (LBL 3/3). Standard deviations of both tests, with scoring and stopping rule variants, are plotted in Figure 2 as a function of the misreporting rate—again, this is the rate at which respondents err, even when they can see and are capable of identifying letters perfectly. Misreports have nothing to do with the respondent’s vision and may be due to blinks, attentional lapses, or any other behavioral factor that is unrelated to the stimulus letter presented. If misreports are substantial, the standard deviation increases, making CS measurement unreliable, as demonstrated by Figure 2, and as first shown by Pelli et al.44 who used sensitivity to misreporting as the most important decision criterion for selecting design and scoring features.

All curves in Figure 2, except the one labeled Pelli-Robson analytic, were produced using the Monte Carlo method and the same three-parameter visual performance model adopted by Pelli et al.11 and Elliott et al.46 The analytic curve was computed using the Pelli et al. closed-form equation. The close concordance of this curve with the curve labeled Pelli-Robson MRS validates the Monte Carlo technique by demonstrating that it makes almost identical predictions.

The Mars test is the most accurate of those shown for misreporting rates lower than 0.002. In this range, the Mars score standard deviation is 28% lower than that of the Pelli-Robson w/MRS; 13% lower than that of the LBL 3/3 method; and 6% lower than that of the LBL 2/3 method. Although Pelli et al. conjectured that the misreporting rate would be substantial, approximately 1% for experienced observers, and up to 5% for inexperienced observers, I recently measured the misreporting rate for verbal letter identification empirically, using large, high-contrast Sloan letters (misreporting rates can be estimated using highly visible stimuli), and found that it was negligible (approximately 0.0005; because my subjects read 15,561 consecutive letters each, it is likely that such a task overestimates the misreporting that would occur in a clinical situation, where observers are vigilant and read only a few letters). The relevant range for assessing relative accuracy of the tests and stopping rules shown in Figure 2 is thus well below 0.002.

Figure 3 shows the mean CS test score for the same conditions as Figure 2. At a misreporting rate of 0.002 or less, the Mars and the Pelli-Robson LBL 2/3 both yield scores very similar to the Pelli-Robson w/MRS (which I consider to be the standard against which bias is measured) and to each other. The Mars test has an average bias of −0.01 log unit over this range; that is, it yields scores that are 0.01 log unit lower than the Pelli Robson w/MRS. The Pelli-Robson LBL 3/3 described by Elliott et al.46 results in a higher, but still acceptable, average bias, yielding scores that are 0.03 log unit higher in CS than those of the Pelli-Robson w/MRS (and 0.04 log unit higher than the Mars test). When Mars test scores are to be interpreted against norms collected using the Pelli-Robson w/MRS, the adjustment of 0.01 log unit can be applied. When interpretation against one of the other scoring-stopping variants is needed, the appropriate bias adjustment can be computed or estimated from Figure 3.

DISCUSSION

The results demonstrate that an improved letter CS test can be designed with identical proximal stimuli (Sloan letters) as the Pelli-Robson test, and in the same sized space (eight rows of six letters). Furthermore, the improved test has a simple and unambiguous scoring procedure that results in greater accuracy than the Pelli-Robson method, using either the manufac-
turer-recommended scoring procedure or any of the scoring variants commonly in use. In addition, the new test has a nearly negligible (but calculable) bias relative to Pelli-Robson and its scoring variants, making scores between the two tests and with published norms directly comparable.

The results also showed that accuracy and test score of the Pelli-Robson depend on the scoring procedure and stopping rule used, complicating interpretation, especially for studies that do not report full details of the stopping rule and scoring used. These complications are avoided with the new test, because it incorporates the virtues of letter-by-letter scoring and therefore is unlikely to be improved by the use of scoring modifications such as those suggested by Elliott et al.46

Note, however, that the results reported herein are all based on analytic modeling or Monte Carlo simulations, as were the results of Pelli et al.44 in their original paper describing the design of the Pelli-Robson test, and as were the subsequent scoring improvements described by Elliott et al.46 These computations have the virtue of allowing repeated testing a great many times with a simulated observer with threshold behavior that mimics that exhibited by actual human observers. Ultimately, however, the most important evaluations of the test will be those that empirically assess its validity and reliability in humans. Such studies are currently under way or planned at several research centers, including Dalhousie University’s Department of Ophthalmology and Visual Sciences (Haymes S, personal communication, December 6, 2004), The Ohio State University’s College of Optometry (Bullimore M, personal communication, January 20, 2005), and Moorfield’s Eye Hospital, London (Crossland M, personal communication, August 2, 2004).

Provided that the chart is manufactured in a controlled and precise fashion, the smaller size of the new test may further contribute to increased reliability above and beyond the improvements due to scoring design. This is so for two reasons: It is more likely to be illuminated evenly; and, being storable and protectable from mechanical and light exposure damage, it is less susceptible to physical degradation. Both of these factors tend to further enhance the precision of CS measurements made with the test.

References