The Amblyopia Treatment Study Visual Acuity Testing Protocol

Jonathan M. Holmes, BM, BCh; Roy W. Beck, MD, PhD; Michael X. Repka, MD; David A. Leske, BS; Raymond T. Kraiker, MSPH; R. Clifford Blair, PhD; Pamela S. Moke, MSPH; Eileen E. Birch, PhD; Richard A. Saunders, MD; Richard W. Hertle, MD; Graham E. Quinn, MD; Kurt A. Simons, PhD; Joseph M. Miller, MD; for the Pediatric Eye Disease Investigator Group

Objective: To evaluate the reliability of a new visual acuity testing protocol for children using isolated surrounded HOTV optotypes.

Methods: After initial pilot testing and modification, the protocol was evaluated using the Baylor-Video Acuity Tester (BVAT) to present isolated surrounded HOTV optotypes. At 6 sites, the protocol was evaluated for testability in 178 children aged 2 to 7 years and for reliability in a subset of 88 children. Twenty-eight percent of the 178 children were classified as having amblyopia.

Results: Using the modified protocol, testability ranged from 24% in 2-year-olds to 96% in 5- to 7-year-olds. Test-retest reliability was high (r=0.82), with 93% of retest scores within 0.1 logMAR unit of the initial test score. The 95% confidence interval for an acuity score was calculated to be the score ±0.125 logMAR unit. For a change between 2 acuity scores, the 95% confidence interval was the difference ±0.18 logMAR unit.

Conclusions: The visual acuity protocol had a high level of testability in 3- to 7-year-olds and excellent test-retest reliability. The protocol has been incorporated into the multicenter Amblyopia Treatment Study and has wide potential application for standardizing visual acuity testing in children.


Quantifying visual acuity is an integral part of the clinical management of many eye diseases in children. Visual acuity is also a desirable outcome measure in clinical trials that address pediatric eye disease. In contrast to testing visual acuity in adults, in which the Early Treatment Diabetic Retinopathy Study (ETDRS) testing protocol is the accepted standard for clinical research, there is no such standard for testing visual acuity in children. The ETDRS protocol has been successfully used in children aged 5½ years and older but is too difficult for many younger children to complete. For preverbal children, the Teller acuity card procedure has been developed and validated as a measure of grating acuity. However, for the age range of 3 to 5 years, ages at which optotype testing is possible, no standardized testing method is commonly accepted. Several different picture and optotype charts are available for testing children in this age range, but high testability in younger children and high specificity for amblyopia are rarely found in the same test. Picture tests, such as Allen cards, have a high degree of testability in preschool children but tend to overestimate visual acuity in amblyopia. As with the ETDRS protocol, other tests based on Snellen optotypes, requiring naming of letters, are too difficult for younger children to complete. Matching tasks have a high level of testability in younger children and can incorporate surround bars, or “crowding bars,” which improve detection of visual acuity loss due to amblyopia.

Several studies have evaluated the optimal letters for testing visual acuity in children. The Screening Test for Young Children and Retardates (STYCAR) incorporates 9 letters (HOTVLXAU) and was developed by Sheridan in 1960. Subsequently, subsets of these 9 letters were proposed to increase testability. One related set of 6 optotypes (XOLOYH) has been incorporated into the Glasgow acuity cards, which presents lines of 4 letters surrounded by a crowding box. The simplest form of the STYCAR is the well-known 4-letter HOTV subset, which was described by Lippmann in 1969 and extensively studied by Friendly in 1978.
The HOTV combination has been demonstrated to be a letter subset with high testability when used with a matching card for ages as young as 2 to 3 years. Never-
theless, little work has been done on standardizing the presentation of the HOTV optotypes. Previous studies most often used lines of HOTV optotypes with varying numbers of letters per line. Sometimes a "mask" was used to present each HOTV letter on a line as a single letter.
MODIFICATION OF THE VISUAL ACUITY TESTING PROTOCOL

The acuity protocol was modified to refine the estimate of threshold by adding a “reinforcement phase,” followed by a second testing for threshold. Analogous to part of a “staircase” procedure, and applying methods developed for clinical Teller acuity card testing,21 larger optotypes are presented to “get the child back on track” or “on a roll.” Reinforcement consists of presenting 3 optotypes of descending size, starting 3 levels above the level failed in phase 1. At the discretion of the tester, the reinforcement phase could be repeated after a break of a few minutes (this was rarely used, and after completion of the present study it was omitted from the protocol). The responses from the reinforcement phase do not contribute to the visual acuity score. Regardless of whether the responses in the reinforcement phase are correct, the last logMAR level failed in phase 1 is retested by the same procedure used in phase 1 (called “phase 2”). If this level is again failed, the test is stopped. If the level is passed, the test continues with smaller levels until a level is failed. The final visual acuity is defined as the smallest level passed in phase 1 or phase 2. The entire testing protocol is given in Table 1, and the scoring sheet is shown in Figure 2.

TESTING OF THE MODIFIED PROTOCOL

After informed consent was obtained, the modified protocol was used to test acuity in 178 children at 6 sites. One primary site contributed 92 children (52%) and the remaining 5 secondary sites contributed 86 children (48%); range, 9-26 children per site). Each BVAT was calibrated before the study according to the manufacturer’s instructions. An Internet-based program, simulating the testing protocol on the BVAT, aided the training of the visual acuity testers.21

Children aged 2 to 7 years were recruited without regard to diagnosis. Children with developmental delay were included so that the assessment of testability would reflect the population of patients who present to pediatric eye disease practices. Developmental delay was assessed by simply asking the parent whether the primary caregiver had identified developmental delay. No attempt was made to formally evaluate the degree of developmental delay. At the conclusion of the office visit, each child’s clinical diagnosis was recorded by the treating physician or by a tester in communication with the treating physician.

Each child underwent acuity testing of the right eye first and then the left eye using the described protocol on the BVAT. A child was considered to be “testable” when the protocol could be completed for both the right and left eyes on the first examination. Testable children were then retested by a separate examiner who was unaware of the results of the first test. The second test was typically performed within 5 to 15 minutes of the first test and always during the same office visit. No attempt was made to randomize, or otherwise determine, the tester order at each site.

Frequently, no specific presentation protocol was used. Even when a protocol for linear presentation of HOTV optotypes was adopted,20 younger children were sometimes confused by lines of optotypes, providing a rationale for the common clinical practice of using single optotypes.

Data from all 178 children were used in the testability analysis. However, only data from the 88 children who completed testing of both eyes twice, without tester errors in both examinations, were used in the test-retest reliability analysis. For testable children, each visual acuity test scoring sheet was evaluated to determine whether tester errors had been made that would affect the acuity score. Tester errors occurred more often at the secondary sites (48% of children) than at the primary site (8% of children). Tester errors did not affect the assessment of testability because completion of the test for each eye did not depend on whether the testing protocol was followed precisely. When present, errors were usually due to unfamiliarity with the protocol early in the implementation process. This prompted additional training of the testers, which reduced the frequency of errors. Data from the children tested at the shorter testing distance for optotype sizes larger than 20/125 were included in the testability analysis but not in the test-retest reliability analysis because reliability might vary according to testing distance. Such differences in reliability between testing distances would be expected because of differences in attention and greater potential errors in estimating visual acuity induced by subject movement toward and away from the screen at the shorter distance.

STATISTICAL METHODS

Testability was evaluated for each age year in a frequency distribution. The association of age-adjusted testability with developmental delay, ocular diagnosis, and performance site was assessed using the Cochran-Mantel-Haenszel statistic.

For the test-retest reliability analysis, visual acuity data were converted to logMAR equivalents. Right and left eye data were combined for analysis after determining that results were similar for separate eye-specific analyses (combining eyes is a valid method of analysis provided that statistical tests adjust for the decreased variance produced by the intereye correlation). Frequency distributions of the differences in visual acuity scores between the first and second test in each eye were constructed and Pearson correlation coefficients were computed overall, and in subgroups, based on diagnosis, visual acuity, age, site, and whether the child had previous experience with HOTV testing. Similar analyses were conducted for interocular difference (IOD) data. Statistical comparisons of proportions were made using a χ² test adjusted for the intereye correlation.23 The SE of measurement for the test was computed separately for each eye and then averaged for reporting and establishing a 95% confidence interval (CI) for an acuity score.24 A 95% CI for a change in an acuity score from a baseline value was computed by 2 methods: one based on the SE of measurement from the first test and one based on the SD of the test-retest differences. These were calculated separately for right and left eyes and then averaged. The CI for the IOD was based on the SD of the test-retest difference in IOD.

Statistical analyses were performed using statistical software (SAS PC version 8.01; SAS Institute Inc, Cary, NC).
cal trial comparing part-time patching with atropine therapy in moderate amblyopia, we recognized the need to develop a formal testing protocol for use with the HOTV letters. The purpose of the present study was to develop and refine a visual acuity testing protocol for isolated surrounded HOTV optotypes.

![Figure 1](https://archopht.jamanetwork.com/)

**Figure 1.** The appearance of a single surrounded HOTV letter.

---

### Table 1. Instructions for Visual Acuity Testing Protocol

<table>
<thead>
<tr>
<th><strong>Binocular Pretest (for Children Who Have Never Performed This Test)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Present 20/125 single surrounded optotypes before the child’s eye is patched. If 2 incorrect, restart pretest at 20/400. If 2 incorrect again, the child is defined as untestable using this testing protocol. If 4 of 4 or 5 of 6 can be identified at either level, proceed with testing.</td>
</tr>
</tbody>
</table>

**Screening**

Patch the eye not being tested. Project a 20/100 or 20/400 single surrounded HOTV optotype (depending on the level passed on the pretest and on the expectation of visual acuity based on previous testing). If correct, go down a logMAR level and again show a single optotype. Continue through 20/20 with 1 letter per level until an incorrect response. If incorrect at either 20/100 or 20/80, go to 20/400 and restart screening. Mark correct letters with a circle on the scoring sheet and incorrect letters with an “X.”

**Phase 1**

- Move up 2 letter sizes (levels) from the letter size with the incorrect response in screening up to the maximum size of 20/400† (eg, if 20/25 is missed in screening, move up to 20/40 in phase 1).
  - Present 4 new letters (if first 3 new letters are correct, the fourth does not need to be tested; as soon as 2 letters are missed, testing of a level stops).
  - If less than 3 letters are correct, proceed to the next largest size and so on until 3 of 4 are correct. When 3 of 4 are correct, proceed to “reinforcement.”‡ If 20/400 is failed, stop testing.
  - If 3 letters are correct, repeat on next smallest optotype. Continue to move to smaller optotypes as long as first 3 or 3 of 4 are correct. If 20/15 is passed, test is over. When 2 letters on a level are missed, stop and move to reinforcement.

**Reinforcement**

The purpose is to get the child back on track. Move up 3 levels from the level missed in phase 1 and show 3 successively smaller single letters (eg, if 20/30 is passed in phase 1 and 20/25 failed, move up to 20/50 in reinforcement and show a 20/50 letter, then 20/40, then 20/30). If the patient fails phase 1 at 20/320 or 20/240, show three 20/400 letters but still start phase 2 at the level failed in phase 1. Whether or not all 3 are correct in reinforcement, proceed to phase 2.

**Phase 2**

Retest the last level failed in phase 1. Continue the test by the same procedure as described for phase 1, with the exception that if 2 letters are missed, testing stops (do not proceed to a larger size).

**Recording Visual Acuity**

Acuity is the smallest letter size (level) passed in phase 1 or phase 2 (eg, smallest logMAR level at which at least 3 presentations [3 of 3 or 3 of 4] are correctly identified).

---

*If the patient will not give a response for a letter, encourage him or her to guess. Be sure to mark an “X” to indicate letters that were shown for which the patient would give no response. Specific instructions for using the protocol with a Baylor-Video Acuity Tester (BVAT): 20/70 letter size is not part of logMAR progression and is skipped; because the largest surrounded HOTV optotype is 20/125 when calibrated for a 10- to 20-ft (3-6 m) distance, the testing distance must be reduced to one-fourth for testing of acuity from 20/160 to 20/400.

†If all letters are correct in screening through 20/20, start phase 1 at 20/30.

‡If the last level tested in phase 1 is passed, the reinforcement phase can be skipped at the tester’s discretion.

---

### RESULTS

**Testability Assessment**

The testability of the final protocol was evaluated in 178 children aged 24 to 94 months; 11% were considered to have developmental delay by parental report. Thirty-four percent were classified as having amblyopia or another cause of decreased acuity (Table 2).

Most children aged 3 years or older could be successfully tested using this protocol: 24% (6/25) of the 2-year-olds, 67% (14/21) of the 3-year-olds, 87% (52/60) of the 4-year-olds, and 96% (69/72) of the 5- to 7-year-olds (Figure 3).

Not unexpectedly, the 20 children classified as developmentally delayed had lower testability (55%) compared with the rest of the cohort (82%) (P<.001). Testability was not statistically different among ocular diagnosis categories and among the performance sites.

The low-vision protocol (for testing acuity at levels <20/125) was used in the testing of only 6 children, all of whom were testable.

**Reliability Assessment**

The number of children for the reliability analysis was reduced to 88, excluding those who were not testable (n=37), who were tested using the low-vision protocol (n=6), whose examinations had tester errors in admin-
The IOD test-retest correlation was 0.75. The retest IOD was within 0.1 logMAR unit of the initial test score. In 93% of the 176 eyes, the retest score was within 0.1 logMAR unit of the initial test score. On the initial test, 86% of the children (85%) had test-retest differences in scores of 0.092 logMAR unit (0.1 logMAR unit).

The 95% CI for an acuity score was calculated to be the mean test-retest score ±0.125 logMAR unit (based on the SE of measurement). For a change between 2 acuity scores, the 95% CI was the difference ±0.21 logMAR unit. With our final protocol, test-retest reliability was excellent, with 93% of retests within 0.1 logMAR unit and for a change in IOD to be the difference ±0.21 logMAR unit.

There were no meaningful differences in the results when the data were analyzed separately for right and left eyes (test-retest correlations: 0.84 and 0.81, respectively; test-retest scores within 0.1 logMAR unit: 93% and 92%, respectively; and SE of measurement: 0.060 and 0.067, respectively). There were also no differences when the data were analyzed by subgroups based on site, diagnosis, age, and whether the child had previous experience with HOTV testing (data not shown).

The correlation between initial and retest visual acuity scores was 0.82. In 93% of the 176 eyes, the retest score was within 0.1 logMAR unit of the initial test score. Data plotted as the test-retest difference vs the mean test-retest score suggest that test-retest reliability did not vary with the level of visual acuity. The 95% CI for an acuity score was calculated to be the score ±0.125 logMAR unit (based on the SE of measurement of 0.064). For a change between 2 acuity scores, the 95% CI was the difference ±0.18 logMAR unit (when based on either the SE of measurement or on the SD of the test-retest difference in scores of 0.092).

The IOD test-retest correlation was 0.75. The retest IOD was within 1 logMAR level (0.1 logMAR unit) of the initial IOD in 75 children (85%) (Figure 7). The 95% CI for the IOD was calculated to be the IOD ±0.15 logMAR unit and for a change in IOD to be the difference ±0.21 logMAR unit.

We have described a structured protocol for measuring visual acuity using isolated surrounded HOTV optotypes in children aged 2 to 7 years. After finding that our original protocol had a lower than desirable test-retest reliability, we added the testing of additional optotypes around threshold to further refine its estimate. With our final protocol, test-reliability was high for children aged 4 years or older, moderate for those aged 3 years, and low for those aged 2 years. Test-retest reliability was excellent, with 93% of retests within 0.1 logMAR unit of the initial test score.
In the clinical management of children with amblyopia, considerable weight is given by many clinicians to the IOD. In our study, the test-retest reliability of the IOD was no better than the reliability of the visual acuity measurements themselves, and perhaps less so.

Although we initially had a high rate of tester error in administering the visual acuity protocol, most of these errors occurred at secondary sites by inexperienced testers. This indicated the need to modify the instructions to the testers, which led to fewer tester errors as the study progressed. Our experience in the ongoing Amblyopia Treatment Study, in which the acuity protocol is being used at 52 sites recruiting 400 children, is that with simple instruction, testers learn to follow the protocol quickly, and errors affecting the visual acuity score are few. Ongoing experience with the visual acuity protocol, in our clinical practices, also confirms that the protocol is not too difficult to teach, learn, and implement.

This indicated the need to modify the instructions to the testers, which led to fewer tester errors as the study progressed. Our experience in the ongoing Amblyopia Treatment Study, in which the acuity protocol is being used at 52 sites recruiting 400 children, is that with simple instruction, testers learn to follow the protocol quickly, and errors affecting the visual acuity score are few. Ongoing experience with the visual acuity protocol, in our clinical practices, also confirms that the protocol is not too difficult to teach, learn, and implement.

Using a separate tester for the retest added a second source of potential variability such that the variability due to the child and that due to the tester cannot be separated. However, in designing the study we believed that this was outweighed by the value of having the second tester masked to the result of the first test. It seemed probable that most of the variability is related to the child’s performance rather than to the tester.
Although we published methods of measuring visual acuity in children, only 16 children.

Our addition of a second attempt at "the level failed" on the initial threshold testing increases the probability that a child will have an acuity score better than threshold strictly by guessing. The probability of achieving a score 1 logMAR level (0.1 logMAR unit) better than threshold by guessing could be as high as 9.9%, but the probability of achieving a score 2 or more logMAR levels (0.2 logMAR unit) better than threshold by guessing is less than 1%. However, the exact probabilities under testing conditions cannot be calculated because they are conditional on whether the acuity threshold is reached during testing and depend on where between 2 logMAR levels true acuity lies and the shape of the psychometric function. This increase in the probability of having an acuity score 1 logMAR level (0.1 logMAR unit) better than threshold by guessing, with the addition of phase 2, is outweighed by a reduction in the likelihood that the acuity is underestimated because of errors made by the child owing to inattention. Future work is needed to define the shape of the psychometric functions obtained from preschool children using this protocol.

Whereas our protocol is designed for testing acuity of 20/400 to 20/15, our test-retest reliability analysis was principally conducted on eyes with visual acuity of 20/30 or better. In addition, most of our children had no ocular cause for decreased acuity and were aged 4 to 7 years. Nevertheless, neither the data from this study nor our ongoing clinical experience using this protocol on a daily basis suggest that the reliability varies according to age, diagnosis, or level of acuity. Further reliability testing would be needed to formally establish this. We assessed the testability of the low-vision protocol for optotype sizes less than 20/125 in only 6 children.

Our acuity protocol compares favorably with other published methods of measuring visual acuity in children in terms of testability and reliability. Although we are unaware of studies that report testability for isolated surrounded HOTV optotypes, Friendly13 and Kastenbaum et al25 described a pattern of increasing testability by age similar to our results for isolated unsurrounded HOTV optotypes. Friendly13 described 15 (88%) of 17 children younger than 3 years as testable, whereas Kastenbaum et al25 described none of 17 children younger than 41 months as testable. Testability approached 100% after age 3½ years in the study by Friendly and after 5 years for Kastenbaum et al. In a study of line HOTV presentation, for 167 children aged 3 to 5 years, Sprague et al20 found that testability ranged from 84% in 3-year-olds to 97% in 5-year-olds. However, 22% of 3-year-olds and 17% of 5-year-olds needed "isolation" of the letters to complete testing. In a study of 105 children aged 2 to 7 years using surrounded "C" and "O" optotypes, Fern and Manny3 found a similar "testability by age" profile to ours, with testability ranging from 30% in 2-year-olds to 100% in those 5 years and older. Regarding reliability, Harvey et al tested line HOTV acuity in 36 children aged 3½ years and 48 aged 4½ years as part of the multicenter Cryotherapy for Retinopathy of Prematurity (CRYO-ROP) study. They reported test-retest rank correlation coefficients of 0.81 and 0.85, respectively, with 40% and 33%, respectively, of retests having exact agreement and approximately 90% within 1 octave (0.3 logMAR unit). For ETDRS testing of 44 children aged 5½ years, Harvey et al found approximately 30% exact agreement and 90% within 1 line (0.1 logMAR level, ie, 0.1 logMAR unit). Although Sprague et al20 reported test-retest reliability for HOTV testing in 3- to 5-year-olds, the results were scored by letter, and he discarded the data from the first one third of the children, so the studies are not comparable. McGraw et al26 studied 68 visually normal children (mean ± SD age, 5.3 ± 1.15 years) using the Glasgow acuity cards incorporating 6 letters (XVOHUY) in lines of 4 surrounded by a crowding box. They found 95% of retests to be within 1 logMAR level (0.1 logMAR unit). Studies of adults using the ETDRS chart show simi-

![Figure 6. Bland-Altman plot of test-retest difference vs mean test-retest visual acuity in 176 eyes of 88 children. The shaded area indicates the 93% of test-retest difference scores that are 0.1 logMAR unit or less (≤ 1 logMAR line) from the 7% of scores that are greater than or equal to 0.2 logMAR unit.](https://archophth.jamanetwork.com/)

![Figure 7. Distribution of test-retest differences in interocular difference (IOD) scores using the final protocol in 88 children. The IODs for the initial test and the retest were calculated as the logMAR acuity score in the right eye minus the score in the left eye. The difference in IOD was then calculated as the IOD from the initial test minus the IOD from the retest.](https://archophth.jamanetwork.com/)
lar results; Elliott and Sheridan27 studied 29 patients with cataract and found that 95% of retest visual acuity scores differed by no more than 0.1 logMAR unit.

We believe that our final protocol is a reasonable compromise between testability and precision for acuity testing of children, using current technology in the preschool age range, in usual office practice and in multicenter clinical trials. Increasing the total number of letter presentations would be problematic in the testing of many young children due to their limited attention span. Under ideal circumstances, given a great deal of time for each patient and excellent patient cooperation and attention, a formal staircase procedure or similar type of testing strategy might be used to estimate the visual acuity threshold, with several reversals around the estimated threshold.28

Alternatively, acuity could be measured by testing multiple optotypes at every logMAR level, similar to the ETDRS protocol.2 Such strategies might address a weakness of our protocol, that of measuring visual acuity when the response on a single letter determines whether the acuity is assigned as 1 logMAR level or the next.

We are currently developing a PC version of our acuity protocol to present HOTV optotypes in a logMAR progression from 20/400 to 20/16. This will eliminate the difficulties in using the BVAT at 2 test distances to measure acuity lower than 20/125 with surrounded HOTV optotypes. It will also reduce the training necessary for testers in the implementation of the protocol and minimize tester errors. We are also developing an alternative PC-based testing strategy using a modified binary search.
approach that holds promise for increasing precision and providing a measure of variability of the point estimate of visual acuity threshold without reducing testability. The PC approach will also readily permit use of different character sets, which can be selected based on the child’s age and development. Incorporating more than 4 letters will reduce the probability of passing a level by chance, although this must be balanced against the possibility that it will reduce the testability of younger children. Adding even just 1 letter so that there are 5 possibilities (eg, HOTVX) substantially reduces the chance probability of passing a logMAR level.

In summary, we developed a new visual acuity testing protocol for isolated surrounded HOTV optotype presentations. The protocol shows good testability in 3- to 7-year-olds and has excellent test-retest reliability. Although we recognize that some of the older children might perform ETDRS testing, maintaining the same testing method is important in a multicenter, longitudinal study because visual acuity test performance can vary according to the nature of the optotype presented. Our acuity protocol provides standardized visual acuity testing in children and has been incorporated into the ongoing Amblyopia Treatment Study.

Accepted for publication February 12, 2001.

This study was supported by cooperative agreement EY11751 from the National Eye Institute, Bethesda, Md, and by Research to Prevent Blindness Inc, New York, NY.

Corresponding author: Jonathan M. Holmes, BM, BCh, Mayo Clinic, Ophthalmology West 7, Rochester, MN 55905 (e-mail: holmes.jonathan@mayo.edu).

Reprints: PEDIG Data Coordinating Center, Jaeb Center for Health Research, 3010 E 138th Ave, Suite 9, Tampa, FL 33613.

REFERENCES