Assessment of visual acuity in infants and young children is a two-step process in which the child responds to a visual stimulus either behaviorally or verbally, and an acuity tester interprets the child's behavioral or verbal response as indicating whether the child can see the stimulus. For example, in the Teller Acuity Card (TAC) procedure, the tester relies on a child's preferential looking behavior to determine whether the child can see striped patterns of varying spatial frequency. In measuring letter acuity, the tester must determine whether a child who does not identify any more letters has reached his or her acuity threshold or whether he or she is no longer attending to the task. Because acuity tests with infants and young children rely on the skill and judgment of the tester and the behavior of the subject, it is important to examine reliability between testers. It is also important to determine whether reliability between observers varies as a function of various patient characteristics, such as age and ophthalmic condition. In a clinical setting, good interobserver reliability is particularly important for consistent assessment of a patient over time, because it is likely that more than one clinician will test a patient's acuity over the course of treatment.

In the multicenter study of Cryotherapy for Retinopathy of Prematurity (CRYO-ROP), visual acuity was assessed at 1- to 1.5-year intervals in more than 1000 children with birth weights less than 1251 g. Grating acuity was assessed in children aged 1, 2, 3.5, 4.5, and 5.5 years, and letter acuity was assessed when children reached ages 3.5, 4.5, and 5.5 years. In most cases, each child's acuity was assessed by only one tester. However, a subset of children, all of whom had severe ROP during early infancy, had either grating acuity or letter acuity assessed by two of the CRYO-ROP study's visual acuity testers.

The purpose of this article is to present interobserver test-retest reliability data for the three visual acuity tests used in the CRYO-ROP study: the TAC procedure,1,23 the crowded HOTV test of letter acuity,4 and the Early-Treatment for Diabetic Retinopathy Study (ETDRS) test of letter acuity.5 For each type of acuity (grating acuity and letter acuity), the influence of age and of severity of retinal residua of ROP on interobserver agreement is examined.
Methods

Subjects

Subjects were preterm children born between January 1986 and November 1987, with birth weights 1250 g or less. All were participants in the randomized portion of the 23-center CRYO-ROP study, in which infants who had severe (threshold) ROP in both eyes had one eye randomly assigned to cryotherapy and the other eye assigned to serve as a control, and infants who had severe ROP in one eye had that eye randomly assigned to cryotherapy or control. Subjects who provided data on interobserver agreement for the present article were located at one of the five study centers at which a CRYO-ROP visual acuity tester was based (Columbus, OH; Minneapolis, MN, Philadelphia, PA; Portland, OR; and Upstate New York [Rochester, Buffalo, Syracuse]) and came in for a CRYO-ROP follow-up examination on a day when two testers (the home tester and a visiting tester) were present.

There were 83 subjects (156 randomized eyes) at the five study centers at which a visual acuity tester was based, and 73 of these subjects participated in interobserver agreement assessment at one or more test ages. Table 1 summarizes the number of interobserver tests completed for each test at each test age. At the 1-year follow-up visit, interobserver tests were conducted on only a small number of eyes, because visual acuity testing was not added to the CRYO-ROP study until after many subjects had completed their 1-year follow-up visit. Data on interobserver agreement for acuity results were obtained on 22 eyes (14%) at 1 year, 80 eyes (51%) at 2 years, 113 eyes (72%) at 3.5 years, 108 eyes (69%) at 4.5 years, and 114 eyes (73%) at 5.5 years. For some eyes, interobserver tests were not completed because the child was lost to follow-up or because the secondary (home) tester was not available. In other instances, interobserver tests were not completed because the child was exempt from acuity testing. To be exempt from acuity testing at a particular age, the child had to have completed the eye examination at that age, and all the following three criteria had to be met: The examining study ophthalmologist found no evidence of visual attention at that age; the examining study ophthalmologist documented that Reese grade 5 ROP was present in both eyes at that age, and the parents agreed that the child showed no evidence of vision at that age. In cases in which the eye examination was conducted on a different day from visual acuity testing, this saved the parents from having to take the child to the study center a second time.

Many participants in this study were visually impaired or developmentally delayed. As is shown in Table 2, approximately one third of the eyes were judged to be blind, and the percentage with acuity within the normal range varied from 25.7% at 3.5 years to 68.2% at 1 year. The decrease after age 1 year in the percentage of eyes with acuity in the normal range has been reported previously and is caused by the failure of many of the eyes to show the improvement in grating visual acuity that occurs between ages 1 and 5.5 years. Table 2 also provides a summary of the developmental status of subjects at each test age who participated in interobserver tests. At the 1-, 2-, 3.5-, and 4.5-year examinations, developmental screening consisted of three age-appropriate questions from the Denver Developmental Screening Test. Areas screened at 1 year were sitting, grasping objects, and responding to sound; at 2 years were feeding, walking, and saying words; at 3.5 years were dressing, drawing, and naming self; and at 4.5 years were placing objects in specific locations, balancing on one foot, and showing understanding of cold, hungry, and tired. A more comprehensive screening, the Functional Independence Measure for Children (WeeFIM) was used to evaluate developmental status at the 5.5-year examination. As shown in Table 2, a substantial number of children failed the developmental screening questions at ages after 1 year.

Visual Acuity Testers

Seven visual acuity testers participated in the CRYO-ROP study. Each primary tester was assigned to test all randomized subjects at five to eight remote study centers, and all subjects in the nonrandomized portion of the study at their home center. To ensure that testers remained masked to the treatment status of each eye of randomized subjects, testers were reassigned to different remote study centers every 6 months and were never assigned to be the primary tester for randomized subjects at their home study centers.

The four testers who were based in Columbus, Philadelphia, Portland, and Upstate New York participated in the study and were trained to use the same visual acuity tester at each study center.
from October 1987 through May 1994. These four testers conducted all acuity testing at the 1-year follow-up examinations, serving as primary testers for randomized patients at remote study centers and as secondary (quality control) testers for randomized patients at their home centers.

For the 2-, 3.5-, 4.5-, and 5.5-year examinations, the number of primary testers was reduced to three (Columbus, Philadelphia, and Upstate NY), but the fourth tester continued to conduct quality control tests and tests of nonrandomized patients at the Portland center. A fifth visual acuity tester, located at the Minneapolis center, began testing in August 1989. The tester at the Minneapolis center was replaced by a sixth visual acuity tester who began testing in January 1990, who was replaced by the seventh visual acuity tester who began testing in February 1992. The Minneapolis testers conducted visual acuity assessments of nonrandomized patients and quality control assessments of randomized patients at the 2-, 3.5-, 4.5-, and 5.5-year follow-up visits but did not serve as the primary tester for randomized subjects.

All testers were trained in the same manner. Initially, testers read the TAC Handbook,15 the CRYO-ROP Manual of Procedures for visual acuity testing,13 and a review of the validation of the TAC procedure.1 The testers then attended a 1-day training session during which they first watched two experienced testers conduct acuity card testing and then tested five to seven infants while the experienced testers watched and instructed them in the specifics of testing. After returning to her home center, each tester tested at least 8 nonstudy infants without ocular abnormalities, and 2 to 10 nonstudy infants with ocular abnormalities. Results were reviewed by one of the trainers, to ensure that acuity results were in the expected range. All testers completed the required testing of nonstudy infants within 6 weeks of the initial training session. Two months after the initial training, testers participated in a 1-day quality control session, in which each tester observed the other testers conduct acuity assessment of nonstudy infants. In addition, testers reviewed testing-related issues with the experimental psychologist (VD) and the pediatric ophthalmologist (GEQ) who were overseeing the acuity assessment protocol.

A similar protocol was followed for training testers in the two letter acuity tests: the crowded HOTV test (Good-Lite, Forest Park, IL),4 which was conducted at the 3.5- and 4.5-year examinations, and the ETDRS letter chart (Lighthouse, Long Island City, NY),5 which was used at the 5.5-year examination.

One-day quality control visual acuity test sessions were held every 6 months between the initial training session in September 1987 and September 1991. During the last 2.5 years of the project, quality control sessions were held at yearly intervals.

### Apparatus
Grating acuity was assessed using a set of Teller acuity cards and a TAC stage (Vistech, Dayton, OH). The set of acuity cards consists of sixteen 25.4 × 56-cm gray cards. Fourteen cards contain a 12.7 × 12.7-cm patch of grating (black-and-white stripes) located to the left or right of a central 0.4-cm-diameter peephole. The gratings range in spatial frequency from 0.32 to 38 cycles/cm in half-octave steps. Also included in the set of acuity cards is a blank gray card and the Low Vision (LV) card, which contains a 22.5 × 25-cm patch of 2.2-cm-wide black-and-white stripes.

Letter acuity was assessed using the crowded HOTV chart or crowded HOTV pocket cards (Good-Lite) at the 3.5- and 4.5-year follow-up examinations. Crowding bars located to the left and right of letters on the HOTV chart and pocket cards reduce the overestimation of acuity that occurs when patients are tested with isolated letters.16,17 At the 5.5-year examination, letter acuity was assessed with the ETDRS log minimum angle of resolution (logMAR) chart (Lighthouse).
Procedures

When the home tester was available, all cooperative randomized subjects at the study centers at which a visual acuity tester was based were tested twice, first by the primary (visiting) tester, whose data were used as primary outcome data in the CRYO-ROP trial, and then by the secondary (home) tester. The primary tester measured the acuity of the right eye first, followed by assessment of the left eye’s acuity. The secondary tester began testing with the left eye, to minimize the number of times the eye patch had to be removed and reapplied.

At the 1- and 2-year follow-up examinations, both the primary and secondary testers measured each subject’s monocular grating acuity. At the 3.5-, 4.5-, and 5.5-year follow-up examinations, the primary tester measured both monocular letter acuity and monocular grating acuity of each subject, if the subject was developmentally able to perform the letter acuity task. Testing of both grating and letter acuity by the primary tester were completed before any testing by the secondary tester. Half of the subjects were randomly assigned to grating acuity assessment by the secondary tester, and half were randomly assigned to letter acuity assessment by the secondary tester. If a subject was assigned to have letter acuity measured by the secondary tester but was too neurodevelopmentally delayed or visually impaired to perform letter acuity testing, the secondary tester instead measured the child’s grating acuity.

Primary and secondary testing was conducted on the same day in 91% of the TAC test-retest comparisons, 97% of the HOTV test-retest comparisons, and 96% of the ETDRS test-retest comparisons. The secondary control tester was unaware of the results of the primary tester, and neither tester had knowledge of which, if either, eye of a subject had received cryotherapy.

Grating Acuity. Monocular grating acuity was measured with the TAC procedure at all five test ages. Standard test distance was 38 cm for 1-year-olds, 55 cm for 2-year-olds, and 84 cm for 3.5-, 4.5-, and 5.5-year-olds. Luminance was 10 cd/m² or more. The acuity score was the spatial frequency of the finest grating that the tester judged that the child could resolve.

Modified (nonstandard) procedures were used with some children. For children with nystagmus, acuity cards were presented vertically, without the acuity card stage. If a child was unresponsive to the lowest spatial frequency grating, the tester reduced the distance between the card and the child (e.g., 84 cm was reduced to 55 cm, then to 38 cm, then to 19 cm, and finally to 10 cm) until a response was observed. If a child provided no evidence that he or she could resolve the 0.52-cycle/cm grating (i.e., the coarsest grating), then the LV card was used as a final attempt to detect pattern vision. The LV card was presented without the acuity card stage and could be held at any distance and moved at any rate necessary to obtain a response. Luminance was 10 cd/m² or more. Acuity was estimated as the line containing the smallest letters on which the subject could match at least four of six letter presentations.

At the 5.5-year examination, monocular letter acuity was tested with the ETDRS chart using a procedure described previously. Standard test distance was 4 m; however, test distances of 1 m, 0.5 m, and 0.25 m were permitted if necessary to obtain a response. Luminance was 10 cd/m² or more. Acuity was scored as the smallest letter size at which the child could identify three out of five letters correctly.

Children whose visual acuity was too poor to allow them to identify or match letters were tested with the TAC procedure, to determine whether any measurable grating acuity was present, or whether the child could detect the grating on the LV card.

Data Analysis

For sighted eyes, agreement between testers’ results was determined by calculating the difference, in octaves, between scores obtained by the two testers. An octave is a halving or doubling of acuity score and is equivalent to 0.301 log unit, or to three lines on a logMar acuity chart.

For data of all eyes (both sighted and blind), the correlation between acuity results of the primary and secondary testers was calculated using Kendall rank correlation coefficients (tau), a nonparametric correlation analysis that includes a correction for tied scores. Categorical agreement between data of the primary and secondary testers was determined by calculation of the kappa statistic. For this analysis, acuity scores from the primary and secondary testers were classified as normal, below normal, poor, or blind, according to the definitions used in the CRYO-ROP study, as shown in Table 3. Blind eyes were those judged to have light perception only or no light perception and those eyes in which the only evidence of pattern vision was detection of the stripes on the LV card. These latter eyes were included in the blind category. Results obtained with the LV card were not quantifiable, because the card could be held at any distance and moved at any rate needed to determine pattern vision. Results for eyes judged to be blind by both testers were included with the results of the type of test (grating or letter acuity) to which they were assigned for assessment of interobserver agreement. However, if an eye was assigned to have interobserver agreement for letter acuity tested, and one or both testers judged that the eye had no measurable letter acuity but had measurable grating acuity, results for that eye were included with the interobserver agreement results for assessment of grating acuity.

In addition, interobserver agreement was determined for eyes categorized in terms of presence and severity of retinal residua of ROP (normal posterior pole, macular heterotomy, and partial or total retinal detachment), as determined by ratings of posterior pole status provided by study ophtalmologists at each test age.

RESULTS

Figure 1 compares acuity results obtained by the primary and secondary testers. Figure 1A provides data for all eyes in which grating acuity was assessed by two testers, and Figure 1B provides data for all eyes in which two testers measured letter acuity with the HOTV (circles) or the ETDRS (triangles) test. Across all ages, interobserver agreement for assessment of
Below normal acuity was selected to be approximately 2 SD below the mean acuity for age. The lower cutoff for normal acuity was 1 octave below the normal range for age, or approximately 3.5 SD below the mean acuity for age. The category of poor acuity included all quantifiable acuity scores below the cutoff for the below normal category. At the time assessment of letter acuity was added to the testing protocol, it was decided to use traditional definitions of normal (20/50 or better at 35 years and 20/40 or better at 4.5 and 55 years) and poor (20/200 or worse) acuity. Thus, categories for grating acuity and letter acuity values may not be entirely comparable. The relation between grating and HOTV acuity results in the CRYO-ROP population has been reported previously.19

As shown in Table 4, Kendall rank order correlation coefficients were 0.75 or better at all test ages, and kappa statistics were 0.68 or better at all ages. Across the five age groups (n = 293), the Kendall rank order correlation coefficient was 0.86, and the kappa statistic was 0.73.

**Letter Acuity.** Figure 3 summarizes HOTV interobserver agreement at 3.5 and 4.5 years and ETDRS interobserver agreement at 5.5 years, for sighted eyes and for eyes judged to be blind by at least one tester. For HOTV data, 62% of interobserver test pairs showed agreement of 0.5 octave or better, and 90% showed agreement of 1.0 octave or better. For ETDRS results at 5.5 years, 90% of interobserver test pairs showed agreement of 0.5 octave or better, and 100% showed agreement of 1.0 octave or better. Of the 38 eyes judged to be blind by one tester, 37 were judged to be blind by the other tester, and one was judged to have acuity in the low vision range (i.e., poor).

As shown in Table 4, Kendall rank correlation coefficients for assessment of letter acuity were 0.81 or better at all three test ages. Kappa values were 0.76 or better at all three ages.

**Effect of Retinal Residua of ROP**

Figure 4 summarizes interobserver agreement for grating acuity (1, 2, 3.5, 4.5, and 5.5 years) and letter acuity (3.5, 4.5, and
Figure 1. Scatterplots of acuity results for primary (x-axis) and secondary testers (y-axis). (A) Grating acuity scores across all five test ages. (B) HOTV (circles) and ETDRS (triangles) letter acuity scores at the 3.5-, 4.5-, and 5.5-year test ages.
FIGURE 2. Interobserver agreement for grating acuity tests. Bars at the left show the percentage of grating acuity interobserver test pairs that differed by 0 octave, more than 0 to 0.5, and more than 0.5 to 1 octave at each of the five test ages for eyes judged to be sighted by both testers. For eyes judged to be blind by at least one tester, the bar at the right shows the percentage of interobserver test pairs in which both testers judged the eye to be blind and the percentage of test pairs in which one tester judged the eye to be blind and the other tester judged the eye to be sighted. Numbers above each bar indicate the number of test pairs represented by the bar. An additional one subject at 1 year, six subjects at 2 years, one subject at 3.5 years, and two subjects at 5.5 years were too uncooperative to complete both tests of grating acuity.

5.5 years) across age groups for sighted eyes with varying degrees of retinal residua of ROP. Interobserver agreement for grating acuity was within 0.5 octave in 63%, 45%, and 36% of test-retest comparisons, and within 1.0 octave in 88%, 76%, and 73% of test-retest comparisons, for eyes with normal posterior pole, macular heterotopia, and partial or total retinal detachment, respectively. Interobserver agreement for letter acuity was within 0.5 octave in 74%, 60%, and 100% of test-retest comparisons and within 1.0 octave in 96%, 85%, and 100% of test-retest comparisons, for eyes with normal posterior pole, macular heterotopia, and partial or total retinal detachment, respectively.

DISCUSSION

The present study provides data on interobserver reliability for measurement of grating acuity at 1, 2, 3.5, 4.5, and 5.5 years and letter acuity at 3.5, 4.5, and 5.5 years in children in the CRYO-ROP trial who were diagnosed with severe ROP. The results indicate that there was excellent interobserver reliability for tests of grating acuity and letter acuity at all ages, and that interobserver reliability was similar for eyes with normal posterior poles, macular heterotopia, and partial or total retinal detachment.

Comparison with Previous Reports

Grating Acuity. Table 5 summarizes interobserver reliability for monocular tests of grating acuity from previous studies and from the present study. Because previous studies included only subjects 4 years of age or younger, we summarized the data from the present study separately for subjects younger than 4 years (1-, 2-, and 3.5-year-olds) and subjects older than 4 years (4.5- and 5.5-year-olds). Interobserver reliability was similar for younger and older test ages in the present study.

Subjects in previous studies included both normal, healthy
TABLE 4. Statistical Analyses Comparing Acuity Results of the Primary and Secondary Testers

<table>
<thead>
<tr>
<th>Acuity Test</th>
<th>Test Age (y)</th>
<th>n</th>
<th>Categorical Agreement (kappa)</th>
<th>Kendall Rank Order Correlation Coefficient (tau)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grating Acuity</td>
<td>1</td>
<td>21</td>
<td>0.90</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>74</td>
<td>0.68</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>73</td>
<td>0.70</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>60</td>
<td>0.75</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>65</td>
<td>0.68</td>
<td>0.89</td>
</tr>
<tr>
<td>HOTV</td>
<td>3.5</td>
<td>36</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>4.5</td>
<td>48</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>ETDTRS</td>
<td>5.5</td>
<td>44</td>
<td>0.84</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Letter Acuity. Table 6 summarizes previous studies of test-retest reliability of tests of letter acuity along with the results from the present study. The upper portion of the table summarizes studies of HOTV tests of letter acuity. Our correlation results for HOTV testing were comparable to those reported by Kastenbaum et al. In contrast, Sprague et al. reported a relatively low correlation but good agreement (98%) within 1 line of the HOTV chart. Differences in scoring meth-
FIGURE 4. Interobserver agreement for grating acuity tests and letter acuity tests for sighted eyes with normal posterior pole, macular heterotopia, and partial or total retinal detachment, as judged by a study ophthalmologist at the time of the acuity test. Details are described in Figure 2. An additional sighted subject had no available data on retinal status. An additional six subjects with a normal posterior pole and four subjects with macular heterotopia did not complete grating acuity testing, and an additional one subject with a normal posterior pole, two subjects with macular heterotopia, and two subjects with retinal detachment did not complete letter acuity testing.

Grating Acuity Versus Letter Acuity Interobserver Reliability

Both grating acuity and letter acuity measures showed substantial interobserver test-retest reliability. However, direct comparison between the grating acuity reliability data presented in Figure 2 and the letter acuity reliability data presented in Figure 3 would be inappropriate, because the subset of children who provided the data shown in Figure 2 is not equivalent, in terms of neurodevelopmental abilities, to the subset of children who provided the data shown in Figure 3. As indicated in the Methods section, each child was randomly assigned to participate in either grating acuity or letter acuity interobserver reliability testing. However, if a child assigned to interobserver

Table 5. Reports of Interobserver Agreement for Monocular Tests of Grating Acuity

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Age (M)</th>
<th>n†</th>
<th>0.0 Oct Diff (%)</th>
<th>≤0.5 Oct Diff (%)</th>
<th>≤1.0 Oct Diff (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonald et al.²⁴</td>
<td>Normal</td>
<td>1-12</td>
<td>66</td>
<td>NR</td>
<td>NR</td>
<td>86</td>
</tr>
<tr>
<td>McDonald et al.²⁵</td>
<td>Normal</td>
<td>18-36</td>
<td>72</td>
<td>NR</td>
<td>88</td>
<td>NR</td>
</tr>
<tr>
<td>Mayer et al.²⁶</td>
<td>Normal</td>
<td>1-48</td>
<td>460</td>
<td>45</td>
<td>90</td>
<td>98</td>
</tr>
<tr>
<td>Getz et al.²⁷</td>
<td>Healthy Preterm</td>
<td>3-38</td>
<td>58</td>
<td>34</td>
<td>79</td>
<td>95</td>
</tr>
<tr>
<td>Preston et al.²⁸</td>
<td>Visual abnorm</td>
<td>2-8</td>
<td>40</td>
<td>NR</td>
<td>88</td>
<td>95</td>
</tr>
<tr>
<td>Dobson et al.²¹</td>
<td>NICU-treated</td>
<td>4-12</td>
<td>382</td>
<td>28</td>
<td>58</td>
<td>80</td>
</tr>
<tr>
<td>Dobson and Carpenter²²</td>
<td>NICU-treated</td>
<td>0-24</td>
<td>1020</td>
<td>NR</td>
<td>NR</td>
<td>85</td>
</tr>
<tr>
<td>Mash et al.²³</td>
<td>NICU-treated</td>
<td>4-48</td>
<td>1918</td>
<td>28</td>
<td>67</td>
<td>87</td>
</tr>
<tr>
<td>Getz et al.²⁷</td>
<td>Oc-neuro abn</td>
<td>3-38</td>
<td>58</td>
<td>26</td>
<td>78</td>
<td>91</td>
</tr>
<tr>
<td>Present study</td>
<td>Severe ROP</td>
<td>12-42</td>
<td>109</td>
<td>36</td>
<td>59</td>
<td>83</td>
</tr>
<tr>
<td>Present study</td>
<td>Severe ROP</td>
<td>54-66</td>
<td>79</td>
<td>33</td>
<td>56</td>
<td>87</td>
</tr>
</tbody>
</table>

NR, not reported; visual abnorm, visual abnormalities; NICU-treated, treated in the neonatal intensive care unit; oc-neuro abn, ocular or neurologic abnormalities.

* Dobson and Carpenter²² subjects were tested binocularly at term and monocularly at 4, 9, 11, 17, and 24 months.
† Number of interobserver test pairs.
TABLE 6. Reports of Test-Retest Agreement for Tests of Letter Acuity

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample (No. of Test Pairs)</th>
<th>Test Chart</th>
<th>Scoring Method</th>
<th>Correlation Coefficient</th>
<th>% Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOTV</td>
<td></td>
<td>HOTV</td>
<td>Smallest line with ≤2 wrong</td>
<td>$r_p = 0.90$</td>
<td>NR</td>
</tr>
<tr>
<td>Kastenbaum et al.</td>
<td>2-5 year-olds (35)</td>
<td>HOTV</td>
<td>Total correct letters read</td>
<td>$r_p = 0.40$</td>
<td>97.7% within 1 line*</td>
</tr>
<tr>
<td>Sprague et al.</td>
<td>3-5 year-olds</td>
<td>HOTV</td>
<td>Smallest line with ≤2 wrong</td>
<td>$k = 0.90$</td>
<td>NR</td>
</tr>
<tr>
<td>Headstart (NR)</td>
<td></td>
<td>HOTV</td>
<td>Largest line with ≤2 wrong</td>
<td>$r = 0.96$</td>
<td>95% within 1 line</td>
</tr>
<tr>
<td>Present study</td>
<td>3.5-6 years severe ROP</td>
<td>HOTV</td>
<td>Smallest line with ≤2 wrong</td>
<td>$r_p = 0.68$</td>
<td>90% within 0.1 logMAR (1 line)</td>
</tr>
<tr>
<td>LogMAR letter</td>
<td></td>
<td>ETDRS</td>
<td>Smallest line with ≤1 wrong</td>
<td>$r_p = 0.90$</td>
<td>NR</td>
</tr>
<tr>
<td>Elliot and Sheridan</td>
<td>Older adults (20),</td>
<td>HotV</td>
<td>Total correct letters × 0.02 logMAR</td>
<td>$r_p = 0.90$</td>
<td>NR</td>
</tr>
<tr>
<td>Sheridan</td>
<td>adult cataract patients (29)</td>
<td></td>
<td></td>
<td>$r_1 = 0.98$</td>
<td>95% within 1 line</td>
</tr>
<tr>
<td>Lovie-Kitchin</td>
<td>Normal 11-74-year-olds</td>
<td>Bailey-Lovie</td>
<td>Kitchin and Bailey, 1981</td>
<td>$r_1 = 0.98$</td>
<td>NR</td>
</tr>
<tr>
<td>Blackhurst et al.</td>
<td>Abnormal vision, adults (164)</td>
<td></td>
<td>MPS scoring 33</td>
<td>$r_1 = 0.99$</td>
<td>87% within 0.1 logMAR (1 line)</td>
</tr>
<tr>
<td>Reeves et al.</td>
<td>Normal/early eye disease, adults (56)</td>
<td></td>
<td>Bailey-Lovie Bailey, 1981</td>
<td>$r_p = 0.86$</td>
<td>84% within 0.1 logMAR (1 line)</td>
</tr>
<tr>
<td>Arditi and Cagenello</td>
<td>Practiced adults (78)</td>
<td>ETDRS</td>
<td>Smallest line with ≤2 wrong</td>
<td>$r_p = 0.68$</td>
<td>90% within 0.1 logMAR (1 line)</td>
</tr>
<tr>
<td>Hawkins</td>
<td>Ocular histoplasmosis and controls (317)</td>
<td>Sloan</td>
<td>Smallest line with ≤1 wrong</td>
<td>$r_p = 0.90$</td>
<td>NR</td>
</tr>
<tr>
<td>Simpson and Regan</td>
<td>Normal 14-57-year-olds</td>
<td>Regan</td>
<td>75% threshold 57</td>
<td>$r = 0.74$</td>
<td>NR</td>
</tr>
<tr>
<td>Woods and Lovie-Kitchin</td>
<td>Low vision (32)</td>
<td>HCV</td>
<td>Bailey-Lovie Bailey, 1981†</td>
<td>$r_1 = 0.90$</td>
<td>NR</td>
</tr>
<tr>
<td>Kheterpal et al.</td>
<td>3-8-year-olds, low vision</td>
<td>ETDRS</td>
<td>Smallest line with ≤2 wrong</td>
<td>$r_p = 0.98$</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Reduced vision (36)</td>
<td></td>
<td></td>
<td>$r_1 = 0.98$</td>
<td>55% within 0.08 logMAR, 100% within 0.2</td>
</tr>
<tr>
<td>Present study</td>
<td>5.5-year-olds, severe ROP</td>
<td>ETDRS</td>
<td>Smallest line with ≤2 wrong</td>
<td>$r_p = 0.84$</td>
<td>90% within 0.1 logMAR (1 line)</td>
</tr>
<tr>
<td></td>
<td>(44)</td>
<td></td>
<td></td>
<td>$r_1 = 0.94$</td>
<td>90% within 0.1 logMAR (1 line)</td>
</tr>
</tbody>
</table>

NR, not reported; $r_p$, Pearson correlation; $r_1$, interclass correlation; $k$, kappa; $\tau$, Kendall's tau; MPS, Macular Photocoagulation Study.

* Six-letter lines of the HOTV chart were used. Percentage represents the % of test-retest pairs that agreed within 6 letters. One-line difference on the HOTV chart is not equivalent to 0.1 logMAR.

† A logMAR acuity score indicating the smallest line at which three or more of the five letters in the row were read correctly was assigned. The logMAR score was then adjusted by adding 0.02 for each letter not read correctly in that row or the preceding row and by subtracting 0.02 for each letter read correctly in the next row.

Testing of letter acuity had neurodevelopmental delay or severe visual impairment and was unable to be tested with the letter acuity procedure, interobserver reliability data were obtained from that child using the grating acuity procedure.

To permit comparison of interobserver reliability for grating versus letter acuity in sighted eyes of children with equivalent neurodevelopmental abilities, data were examined from the subset of children in whom the primary tester was able to obtain both a measurable grating acuity score and a measurable letter acuity score. Although interobserver reliability data for letter acuity were obtained in only half of these children, all had the neurodevelopmental ability to complete both acuity measures. As shown in Figure 5, the proportion of interobserver test-retest pairs showing perfect agreement (0.0 octave difference) was similar across all ages for grating acuity versus letter acuity (41.7% versus 40.0% at 3.5 years, 34.4% versus 25.7% at 4.5 years, and 35.5% versus 33.3% at 5.5 years, for grating acuity versus letter acuity, respectively), as was the proportion of interobserver test-retest pairs showing a difference of no more than 1.0 octave (97.2% versus 88.0% at 3.5 years, 84.4% versus 91.4% at 4.5 years, and 93.5% versus 100% at 5.5 years, for grating acuity versus letter acuity, respectively). Also, at 3.5 and 4.5 years, the proportion of interobserver test-retest pairs showing a difference of no more than 0.5 octave was similar for grating acuity and letter acuity (61.1% for grating acuity versus 68.0% for letter acuity at 3.5 years; 56.3% versus 57.1% at 4.5 years, respectively). However, at 5.5 years a larger percentage of letter acuity interobserver test-retest pairs (90.0%) than grating acuity test-retest pairs (61.3%) showed agreement of 0.5 octave or better. This higher level of interobserver agreement with letter acuity at age 5.5 years may reflect the greater accuracy of acuity measurement that is obtained with a task involving identification of 10 letters (ETDRS chart) than the 4-letter identification task used in
FIGURE 5. Interobserver agreement for grating acuity tests and letter acuity tests for the subset of children in whom the primary tester was able to obtain both a measurable grating acuity score and a measurable letter acuity score. Approximately half of the children were randomly assigned to have grating acuity measured by the second tester, and approximately half were randomly assigned to have letter acuity measured by the second tester. Details are described in Figure 2.

HOTV testing or the two-choice detection task used in grating acuity testing.

CONCLUSIONS

The findings suggest that the TAC, HOTV, and ETDRS methods provided reliable measurements of visual acuity in children in the CRYO-ROP study, regardless of the condition of the eye. These results support the validity of previous findings regarding differences in visual acuity in comparisons across varying degrees of retinal residua of ROP. 9,10

It is important to note that, on behalf of the CRYO-ROP study, considerable effort was made by researchers and testers to ensure strict quality control in acuity testing. Acuity testers initially received extensive training and certification. Subsequently, they participated in periodic quality control training sessions throughout the course of the study. This type of training regimen is not typical in most clinical settings. However, the results of the present study suggest that, with proper attention to training and monitoring of testing, reliable measurements of visual acuity can be obtained for children with severe ROP—that is, children with a wide range of developmental differences and visual capabilities.

Acknowledgments

The authors thank the CRYO-ROP 1- to 5.5-year follow-up visual acuity testers: Patricia Bartholomew (Portland, OR), Janice Evans (Philadelphia, PA), Annette Hambel (Upstate NY), Jackie Mogan (Minneapolis, MN), Lisa Romig (Minneapolis, MN), Laura Trueb (Columbus, OH), and Jane L. Verness (Minneapolis, MN).

References