ORIGINAL ARTICLE

Reliability of the CSV-1000 in Adults and Children

Susan A. Kelly*, Yi Pang[†], and Stephanie Klemencic[‡]

ABSTRACT

Purpose. Test–retest reliability of the CSV-1000 (Vector Vision) has only been reported for one adult sample. We measured the reliability of this instrument in both children and adults and also investigated the effect of changing the examiner on test–retest reliability.

Methods. Test–retest log contrast sensitivity (CS) measurements were obtained for 19 young adults and 15 children by the same examiner. Test–retest log CS data were obtained from 21 young adults with different examiners. Reliability was calculated using the Bland–Altman limits of agreement, the coefficient of repeatability (COR), and the intraclass correlation coefficient.

Results. All three estimates of reliability for the CSV-1000 chart are low for both children and adults using the standard recommended testing protocol. If the test–retest log CS data are obtained from the same examiner then the reliability is improved, but not significantly so.

Conclusions. The reliability of the CSV-1000 is low, even if the same examiner obtains test–retest data. The data indicate that this test is unlikely to be sensitive enough to provide useful information for the clinician as is, but we suggest modifications of the procedure that may significantly increase test reliability.
(Optom Vis Sci 2012;89:1172–1181)

Key Words: contrast sensitivity, reliability, Vector Vision, limits of agreement, intraclass correlation coefficient

Contrast sensitivity (CS) testing has become an important clinical tool in the battery of tests used to characterize patients' vision. Most clinicians now recognize that visual acuity is only one measure of visual function and that this particular measure tests the resolution limit of the visual system. Visual acuity is correlated with how well an observer can detect and process high spatial frequencies. CS, on the other hand, measures how much contrast is required to detect a particular spatial frequency. Typically, contrast thresholds are measured across a range of spatial frequencies, and the reciprocal of these thresholds produces a CS function (for an excellent review of the history and current CS testing options, see Owsley[1]).

It is thought that the visual system is composed of a series of spatial frequency filters, each of which detects and processes a limited range of frequencies that are further processed and ultimately result in a visual percept. Most visual percepts in our everyday environment consist of multiple spatial frequencies; that is, if a given visual image is deconstructed, it would consist of many spa-

tial frequencies, which when added together (with proper phase, amplitude, and orientation), would produce the visual image. For example, the ability to recognize a face requires the detection and processing of a range of spatial frequencies, whereas the recognition of a very small letter relies mainly on detection of high spatial frequencies. Thus, CS testing across a range of spatial frequencies allows the clinician to more fully understand the visual deficits a patient is experiencing even if visual acuity is normal or nearly so. For example, after cataract or refractive surgery, the visual acuity may be 20/20, but aberrations may degrade the quality of vision without affecting spatial resolution. As pointed out by Packer et al.,[2] CS correlates highly with functional aspects of daily living, such as driving difficulty, crash frequency, and postural stability, while the link between functional disability and CS impairment is independent of visual acuity loss.[1] CS is also strongly correlated with reading performance, ambulation mobility, and face recognition.[3]

The measurement of CS across a range of spatial frequencies takes time, from many minutes per eye with computer-based charts to 2 min or less with chart-based systems. It has been suggested that because visual acuity measurements estimate how well the visual system processes high spatial frequencies, one could then

*PhD
[†]PhD, OD
[‡]OD
Illinois College of Optometry, Chicago, Illinois.

just measure CS at the frequency humans are most sensitive to [which is mid-range and usually between 3 and 6 cycles per degree (cpd)] and obtain enough information for most patients.[4] Testing time is significantly reduced if CS is only measured at the adult peak spatial frequency. The Pelli-Robson test was designed with this intention in mind.[5] It is a well-designed test that consists of a series of letter triplets, where each series systematically decreases in contrast, from 100% at the top of the chart to 1% at the bottom in 0.15 log unit steps. All letters subtend 2.86 deg at a 1-m test distance. The subject is instructed to identify each letter in a given triplet and is given credit for identifying two of three letters of a given triplet correctly. Testing continues until the contrast is so low the subject can no longer reliably identify two of three letters correctly. Studies have reported excellent test–retest reliability for this test for both visually normal observers as well as those with ocular disease.[6–10]

The argument has been made that assessment of intermediate and high spatial frequencies can provide enough information for the clinician to assess the function of the patient's visual system, and it can do so quickly, as the entire range of spatial frequencies are not tested. However, there are other authors who point out that it is important to sample sensitivity across the full range of spatial frequencies because losses may exist that are not revealed by testing at only the peak (i.e., intermediate) or high end of the CS function. For example, selective losses in CS at low spatial frequencies can occur in patients with cerebral lesions[11] and is an early sign of malnutrition.[12] Also, after refractive surgery, visual acuity may be markedly improved, but losses in CS at low and intermediate spatial frequencies may explain the poor quality of vision that is sometimes reported. The recent report that action video game play can improve CS across almost all spatial frequencies suggests that eye-care professionals may now be able to treat losses in sensitivity at low and mid-spatial frequencies rather than just losses at high spatial frequencies, as has traditionally been the case.[13–16] In addition, CS measurements obtained across a range of spatial frequencies can more fully monitor the efficacy of a treatment option.

However, the ability of a given CS test to assist in describing and/or monitoring vision quality depends on its accuracy. Accuracy is a measure of a test's validity and its reliability. A test is valid if the test result designed to measure a given variable agrees with the true value of the variable. Reliability refers to the level of agreement between the same measurements taken at different points in time. Tests that lack validity and/or reliability will not be sensitive enough to detect improvements in CS that might accompany cataract or LASIK surgery nor will they be able to detect CS losses that often occur with the development of cataract, age-related macular disease, glaucoma, or other ocular pathologies. In addition, tests with low test–retest reliability will have poor agreement with other tests designed to measure the same variable. In addition, CS measurements must not only be accurate but also be quickly and easily obtained in a clinical setting without prohibitive cost.

There are currently several options for CS measurement, but, unfortunately, none of the available systems that measure CS across a range of spatial frequencies (as opposed to just peak CS) can satisfy all of the aforementioned requirements. Computer-based CS test systems have the greatest potential, which is, as yet, unrealized. They have the potential to minimize inter-examiner variability and to employ criterion-free psychophysical protocols that accurately measure CS thresholds. The difficulty, however, at least at this time, is that the protocols needed to produce reliable data are lengthy, although this may change,[17–19] and the video monitors needed to provide stable and linear changes in contrast at low levels are expensive.

An alternative to a computerized test system is to use one of the readily available and reasonably priced chart-based systems. These tests measure CS at multiple spatial frequencies and require subjects to either discriminate the orientation of the sinewave gratings such as the Vistech (VCTS), its successor, the Functional Acuity Contrast Test (FACT), or to detect the presence of a sinewave grating pattern by indicating its location, such as the Vector Vision CSV-1000 system. There are other chart-based tests, but these are either very similar to the Vistech (such as the Sine-Wave Contrast Test) or the FACT (such as the OPTEC 6500). In our experience, computer-based CS tests can require up to 5 min per eye, whereas the chart-based tests typically require 2 min or less for healthy young adults.

Although the CSV-1000 has been used in a number of clinical studies,[20–22] unlike the Vistech, Pelli-Robson, and other CS tests, only one study has investigated the test–retest reliability of this chart in an adult sample[23]; there were no reports in the literature of test–retest reliability in children when we began our study, although a report on the test–retest reliability between examiners has recently been published.[24] The present study measured test–retest reliability of the CSV-1000 in a sample of young visually normal adults and also in a sample of visually normal children aged between 5 and 12 years. Reliability was calculated using the following metrics: (1) limits of agreement [(LoA); both calculated and visually inspected on the Bland–Altman plot], (2) the coefficient of repeatability (COR), and (3) the intraclass correlation coefficient (ICC). The effect of different examiners on test–retest variability was also examined.

METHODS

Subjects

Adults

Test–retest CS measures were obtained from a sample of 40 visually normal adults [mean age 26.4 years, standard deviation (SD) = 4.7 years, range 22 to 38 years] and 15 children (mean age 7.7 years, SD = 2.02 years, range 5 to 12 years). The purpose of the study and testing protocol were explained before participation in the study, and informed consent was obtained from either the subject or the subject's parent/legal guardian. The testing protocol conformed to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board at the Illinois College of Optometry. All subjects either had had an eye examination within 12 months of the CS testing or were examined by one of the authors (SK) before enrollment in the study. If the potential subject was monocularly correctible to at least 0 logMAR (20/20), free from ocular pathology, strabismus, and amblyopia, they were enrolled in the study. Subjects wore their habitual correction during testing if needed. The 40 adult subjects were divided into two groups, one of which had the same examiner for test–retest, whereas the other had a different examiner for test–retest. Because

TABLE 1.
Patient characteristics

	n	Mean age (range) yr	Best-corrected VA	Average test–retest interval
Adults				
Same examiner	19	24 (21–38)	≤20/20	0.83 months
Different examiner	21	27.9 (22–30)	≤20/20	3.54 months
Combined	40	26.16 (21–38)	≤20/20	2.18 months
Pomerance and Evans ('94)	24	63.9	≤20/40	2.65 months
Children				
Same examiner	15	7 (5–12)	≥20/20	2 weeks

the sample size of the children was small (n = 15), we used the same examiner for both visits. These examiners were the same examiners who obtained the test–retest measurements for adults. The sample size, mean age, minimal acuity required for enrollment, and inter-test duration are listed in Table 1 for adults and children. Similar information is included in Table 1 for the Pomerance and Evans study.[23]

CS Measurements

Monocular CS measurements were obtained for all subjects with the Vector Vision CSV-1000 (Greenville, OH) chart during two separate visits (see Table 1 for average inter-test interval). The test consists of a translucent chart that is rear-illuminated by a tungsten bulb. The unit self-calibrates to produce a mean luminance of 85 cd/m². All subjects were tested at the recommended distance of 8 feet. The CSV-1000 consists of a series of circular achromatic sinewave patches 1.5 inches in diameter. Across each row, there are vertical pairs of circles, one of which contains the sinewave patch while the other is blank but the same space-averaged luminance as the test patch. There are four rows, each corresponding to one of four spatial frequencies; 3, 6, 12, or 18 cpd. When selected, a given spatial frequency is rear-illuminated and the subject is shown a suprathreshold example of the test pattern. Each spatial frequency is presented at eight different contrast levels that systematically decrease from 0.20 to 0.08 in eight columns from left to right. The subject is instructed to indicate whether the given test pattern is located in the top or bottom patch. The manufacturer's instructions indicate that it is important to inform the subjects that there are three possible responses: the sinewave pattern is in the top circle, bottom circle, or neither, the latter meaning the subject cannot see a pattern in either of the patches. The contrast threshold is defined as the contrast of the last column the subject could correctly identify the location of the sinewave patch. The average inter-stimulus drop in contrast is 0.15 log units between steps 2 and 8; the contrast change between step 1 and step 2 is 0.3 log units.

Testing Protocol

Adults

Twenty-one of the 40 adult subjects were tested as part of a larger study involved in the collection of normative data for differ-ent CS tests as a function of age. These "test" or first-visit CS values were all obtained by the same tester. Retest values for these subjects were obtained by four new examiners. These 21 subjects comprised the group where test–retest data were obtained by different examiners. The remaining 19 subjects were both tested and re-tested by these same four examiners, but in this group, both test and retest data were obtained by the same examiner. All five examiners were students at the Illinois College of Optometry who had completed at least 1 year of the program and thus had experience performing psychophysical tests. All examiners were trained by the same instructor as to how to run the CS test. All data were obtained monocularly from the subject's dominant eye. Pilot data indicated that monocular CS functions required about 2 min or less to complete for both children and adults.

Children

Fifteen children were tested and re-tested by the same examiner with the CSV-1000 chart. The testing protocol was the same for children as was described earlier for adults.

Data Analysis

All CS data were converted to log CS for statistical analyses. Test–retest reliability was calculated separately for adults and children. The adult data were analyzed separately depending on whether they were re-tested by the same or different examiners. Test–retest reliability was assessed by three measures: (1) the LoA were calculated with the Bland–Altman technique that also allows one to look at the bias (sometimes called accuracy), which is the average difference between test–retest scores,[25] (2) The COR, which is calculated as 1.96 times the SD of the test–retest differences, and (3) the ICC. All statistical analyses were performed with SPSS (Version 17.0, Chicago, IL). All log CS data and reliability measures were compared with the study by Pomerance and Evans[23] where applicable.

The role of the examiner was investigated in several ways. As aforementioned, reliability was measured for subjects with the same examiner and compared with the reliability measures obtained from subjects tested with two different examiners. In addition, we analyzed the effect of the four different examiners using a 3-way analysis of variance (ANOVA; visit, spatial frequency, examiner) and compared the effect of examiner on the reliability measures described earlier. We also used a 2-way repeated measures ANOVA to compare the differences between visits for subjects with the same vs. different examiners.

RESULTS

The mean log CS and SD for test and retest are listed in Tables 2 (adults) and 3 (children). The maximum log CS that can be obtained on the CSV-1000 is as follows: 2.08 (3 cpd), 2.29 (6 cpd), 1.99 (12 cpd), and 1.55 (18 cpd). The mean log CS was higher for all spatial frequencies in the current study than those reported by Pomerance and Evans[23] (for subjects with same or different examiners). It is likely that the greater sensitivity exhibited by our two adult samples is due to their younger average age, as contrast thresholds typically increase with age.[26] Tables 2 and 3 also list the number of subjects exhibiting a ceiling effect, which indicates that

TABLE 2.

Mean log CS for adults

Adults	Test [mean log CS (SD)] Spatial frequency (cpd)				Retest [mean log CS (SD)] Spatial frequency (cpd)			
	3	6	12	18	3	6	12	18
Same examiner (n = 19)	1.83 (0.14)	1.96 (0.27)	1.75 (0.19)	1.17 (0.23)	1.88 (0.18)	2.0 (0.19)	1.65 (0.24)	1.24 (0.23)
Different examiner (n = 21)	1.84 (0.16)	2.16 (0.17)	1.7 (0.26)	1.2 (0.28)	1.86 (0.23)	2.05 (0.25)	1.76 (0.25)	1.31 (0.25)
Pomerance and Evans ('94)	1.55 (0.16)	1.76 (0.18)	1.49 (0.23)	0.91 (0.3)	1.57 (0.15)	1.84 (0.16)	1.5 (0.15)	0.95 (0.23)
No. reaching max log CS (out of 40)	5 (12.5%)	7 (17.5%)	5 (12.5%)	6 (15%)	14 (35%)	12 (30%)	10 (25%)	13 (32.5%)

TABLE 3.

Mean log CS values for children

Children	Test [mean log CS (SD)] Spatial frequency (cpd)				Retest [mean log CS (SD)] Spatial frequency (cpd)			
	3	6	12	18	3	6	12	18
Same examiner (n = 15)	1.73 (0.25)	2.09 (0.20)	1.66 (0.31)	1.23 (0.25)	1.74 (0.25)	2.07 (0.24)	1.67 (0.24)	1.29 (0.21)
No. reaching max log CS (out of 15)	3 (20%)	6 (40%)	6 (40%)	3 (20%)	2 (13.3%)	6 (40%)	3 (20%)	3 (20%)

subjects have not necessarily reached their maximum sensitivity for a given spatial frequency. For adult subjects, the number reaching a ceiling effect does not vary much with spatial frequency, but note that the number of subjects reaching the maximum sensitivity of the test doubles for each spatial frequency during the retest. In other words, 14.4% of subjects reached maximum sensitivity during the test, but almost a third of the subjects reached maximum sensitivity during the retest (30.6%). Nonetheless, 2-way repeated measures ANOVA indicated that retest log CS scores were not significantly higher (or different) from test log CS scores for subjects tested by the same examiner (F = 0.062, p > 0.05) or different examiners (F = 1.39, p > 0.05).

The summary of children's data in Table 3 indicates that there is a slight increase in log CS at 18 cpd during the retest, but a 2-way repeated measures ANOVA indicates that there is no significant difference in log CS between the two visits at any spatial frequency (F = 0.088, p > 0.05). In agreement with the adult data, the difference in log CS between the two visits (Table 5) is very small. Note the children's data showed no such increase in percentage of subjects reaching maximum sensitivity during the retest; 18 of 60 (30%) did so during the test, whereas 14 of 60 (23.3%) did so during the retest.

Reliability

Adult reliability measures are listed in Table 4 along with the mean and SD of the test–retest differences in log CS. These same measures are listed for children in Table 5. Inspection of Table 4 indicates that the average test–retest differences are very small regardless of whether the examiner is the same or different. These small differences are similar to those reported for adults in the earlier study,[23] with the exception of the difference observed at 18 cpd (same examiner). The major difference between the data obtained from the two studies is the size of the SD of test–retest

differences; those obtained by Pomerance and Evans[23] are about half of what is reported in the present study.

Although the mean CS values obtained by Pomerance and Evans[23] are almost the same as those reported in the present study, reliability measures will significantly differ between the two studies owing to the large discrepancy in the SD of test–retest differences because the LoA and COR are both derived from this value. The COR is calculated by multiplying the SD of the differences between visit times 1.96, whereas the LoA are found by adding and subtracting the COR to the mean test–rest difference for each spatial frequency. As can be seen in Table 4, both the LoA and the COR values are much smaller for the previous study.[23]

Another way to assess the agreement between test and retest is to plot the difference between the two visits (test–retest log CS) vs. the average of the two visits as suggested by Bland and Altman.[25] Figs. 1 and 2 plot the results obtained with adult subjects tested by the same examiner and different examiners, respectively. A similar plot is illustrated for children in Fig. 3. The average test–retest difference (bias) is indicated on each plot in both figures as the solid line and listed in Tables 4 (adults) and 5 (children). The bias is very close to zero except for the 18 cpd condition (adults, same examiner) and slightly negative for all conditions except 12 cpd (adults, same examiner), which indicates there is a slight, but insignificant, practice effect, which is not statistically significant (F = 0.062, p = >0.05).

The magnitude of the LoA and their 95% confidence limits for test–retest log CS values are listed in Table 4. The LoA for the data collected by Pomerance and Evans were calculated from the COR values listed in their Table 4.[23] The LoA obtained from the current study are plotted in Figs. 1 to 3 as the dotted lines. Despite the small average test–retest differences, it is readily apparent in Figs. 1 and 2 that the log CS values for adult subjects can differ by up to about 1 log unit between visit 1 and visit 2. This is an enormous

TABLE 4.
Reliability measures for adults

Spatial frequency (cpd)	Adults Same examiner (n = 19)	Difference in test–retest log CS (mean ± SD) Different examiner (n = 21)	Pomerance and Evans (1994) (n = 24)
3	0.05 (0.2)	0.08 (0.22)	0.02 (0.05)
6	0.04 (0.22)	0.07 (0.23)	0.08 (0.07)
12	−0.10 (0.3)	0.06 (0.32)	0.01 (0.12)
18	−0.06 (0.26)	0.09 (0.39)	0.04 (0.15)
LoA, log CS			
3	0.78	0.86	0.17
6	0.86	0.9	0.3
12	1.16	1.26	0.44
18	1.06	1.52	0.6
LoA (log CS) outliers removed			
3	0.78		
6	0.75		
12	1.00		
18	0.74		
COR			
3	0.39	0.43	0.09
6	0.43	0.45	0.15
12	0.58	0.63	0.23
18	0.51	0.76	0.30
COR outliers removed			
3	0.39		
6	0.37		
12	0.50		
18	0.37		
ICC			
3	0.24	0.36	
6	0.57[a]	0.38	
12	0.07	0.18	
18	0.31	−0.59	
ICC outliers removed			
3	0.42		
6	0.65[a]		
12	0.17		
18	0.56[a]		

[a]ICC is statistically significant.

TABLE 5.
Reliability measures for children

Children Spatial frequency (cpd)	Difference in test–retest log CS [mean (SD)] Same examiner (n = 15)
3	0.01 (0.25)
6	0.02 (0.31)
12	0.01 (0.46)
18	0.06 (0.25)
LoA (log CS)	
3	0.96
6	1.23
12	1.82
18	1.0
COR	
3	0.48
6	0.61
12	0.91
18	0.50
ICC	
3	0.55[a]
6	−0.01
12	−0.44
18	0.38

[a]ICC is statistically significant.

illustrate that the differences in log CS between visits do not seem to be systematically related to the average log CS.

Table 4 also lists the test–retest ICC, which are commonly used to quantify the consistency of ratings made by different subjects but can also be used as a measure of test–retest reliability.[27] Like the more familiar Spearman correlation coefficient that measures agreement between classes, the ICC measures agreement within a class and ranges from 0 to1.0 (although it can be negative if there is more agreement between individual subjects than within repeated measures of the same subject). The ICC values listed in Tables 4 (adults) and 5 (children) were calculated using a 2-way random effects model for absolute agreement. ICC values >0.75 are considered excellent, but none of the values listed in Tables 4 and 5 approach that level. However, it should be noted that there is no clear consensus on what constitutes excellent, good, or poor ICC values.[27]

Outliers

Inspection of Figs. 1 to 3 indicates that various spatial frequencies have outliers, which can be defined as those subjects whose test–retest difference exceeds the range expected to contain 95% of the test–retest differences. The Bland–Altman plots in Fig. 1 show that at 6 cpd, one subject (5.3%) has a test–retest difference that exceeds the LoA; at 12 cpd and 18 cpd, four subjects (21.1%) and one subject (5.3%), respectively, exceed the range defined by the LoA. Subjects tested with different examiners and children also produced outliers, but, in the interests of optimizing reliability, we recalculated the LoA, COR, and ICC without outliers for adults
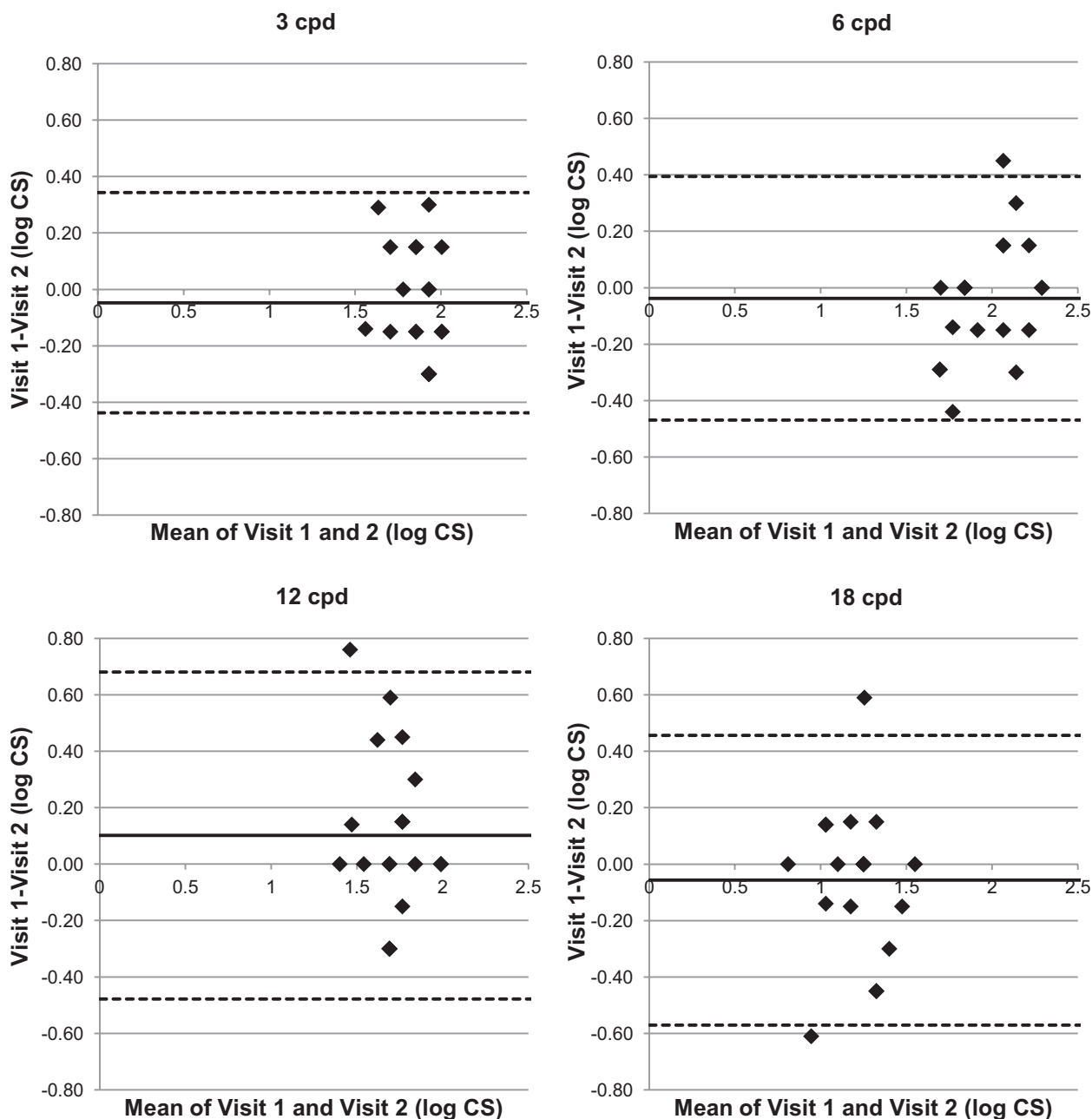
range, which is also evident from inspection of Table 3 and where it is evident that the magnitude of the LoA for the current study is much wider than that reported by Pomerance and Evans.[23] The LoA specify the range of test–retest differences expected to occur in 95% of future subjects similar to those in the present study. Given the large magnitude of this interval for all spatial frequencies, both age-groups, and regardless of whether the examiner is the same or different, the reliability is poor. However, despite the large width of the LoA, the graphs

FIGURE 1.

Bland–Altman plot for adult subjects tested and re-tested with the same examiner (n = 19).

tested by the same examiner only. The recalculated values show a smaller range for the LoA, and improved COR, but these improvements are slight. The ICC values are improved at every spatial frequency, but the only significant improvement occurs at 18 cpd.

Effect of Same vs. Different Examiner

Reliability data from the 19 adult subjects whose test–retest data were acquired by the same examiner are compared with the reliability measures obtained from the 21 adult subjects tested by different examiners in Table 4. Inspection of the data indicates that for every spatial frequency, the LoA and the COR are improved when the test–retest data are acquired by the same examiner. However, the improvements are modest, and the width of the LoA is

still quite large (see Fig. 2) and the COR is still quite low. The ICC values obtained with the same examiner are not uniformly improved in comparison with those obtained with different examiners. We also compared the mean log CS test–retest differences obtained with the same vs. different examiners with a 2-way ANOVA (spatial frequency, examiner). Results indicate no significant difference between the test–retest differences between the two groups for any spatial frequency (F = 1.1, p > 0.05).

DISCUSSION

CS measurements can provide important information that is in addition to, and different from, the information provided by visual acuity. Chart-based CS tests are economical, easy to administer,
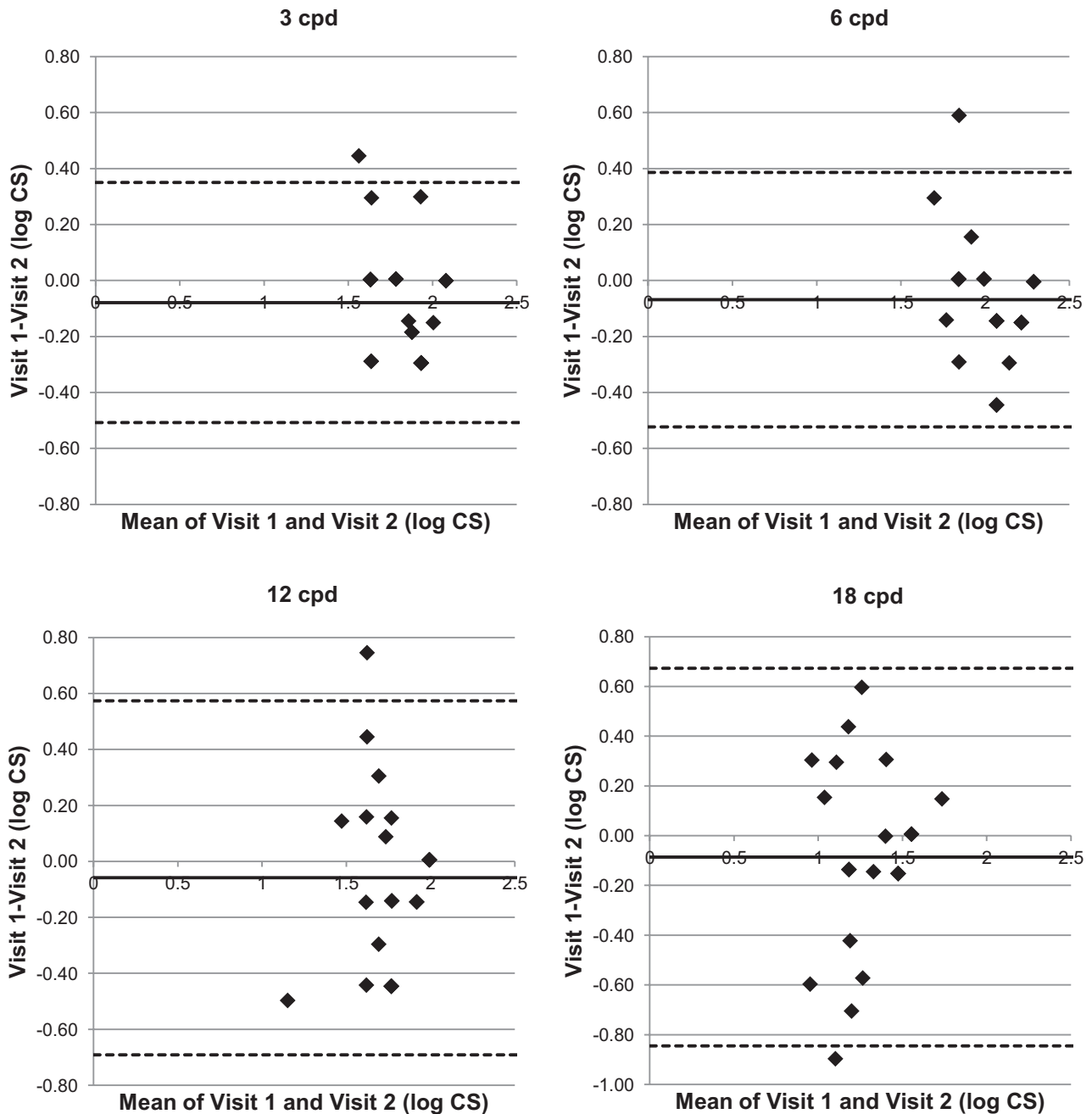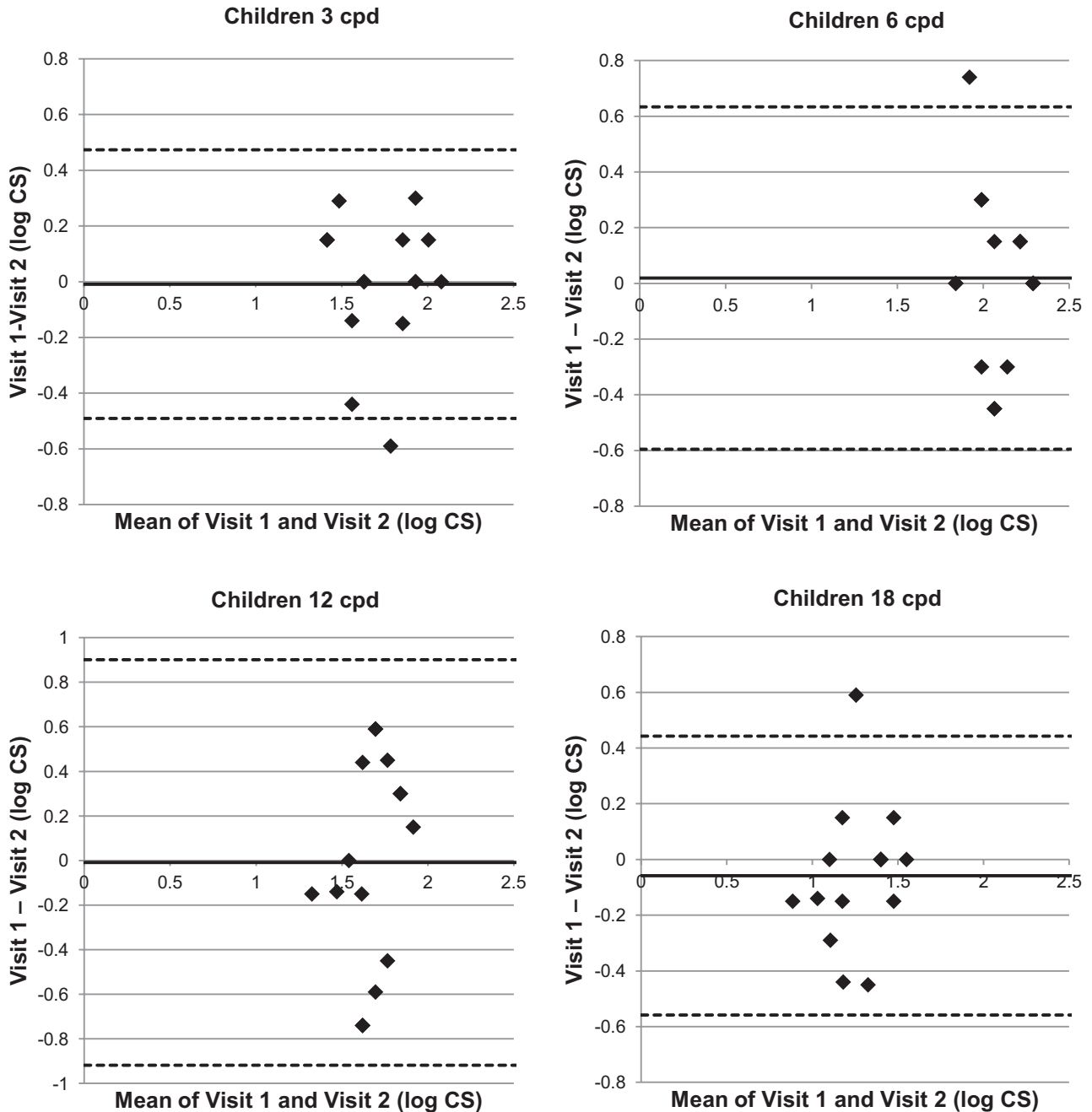
FIGURE 2.

Bland–Altman plot for adult subjects tested and re-tested with different examiners (n = 21).

and take little time; however, even a readily available, easy to administer test will not be useful if test results are not reliable. Our results indicate that according to common measures of reliability, the test–retest reliability for the CSV-1000 is low. As pointed out by Bland and Altman, determination of an acceptable range, as defined by the LoA, is a clinical question, not a statistical one.[25] Although there is no hard and fast rule with respect to CS, a 0.1 log difference in log CS between visits reflects a 25% change, whereas a 0.2 log difference represents a 58% difference in log CS between visits. Clearly LoA equal to a range of 0.1 to 0.2 log units is more desirable than the much larger values listed in Tables 4 and 5.

The LoA as well as the COR listed in Table 4 are much larger than those reported by Pomerance and Evans,[23] even when re-

calculated with outliers removed. It seems consistent with the unreliability of the test that a significant number of healthy young adults would be classified as outliers for no apparent reason: (4 of 19 or 21.05% of those tested with the same examiner and 4 of 21 or 19.05% with different examiners).

Tables 4 and 5 indicate that the statistical measures commonly used to quantify test–retest reliability are not strongly affected by whether the examiner is the same or different. The similarity in reliability measures for same vs. different examiners indicates that the effect of examiner on test–retest differences is small. The analysis of variance performed on the differences between test and retest also yielded no significant difference between the two groups. Pomerance and Evans[23] used different examiners for the

FIGURE 3.

Bland–Altman plot for children. All children tested and re-tested with the same examiner (n = 15).

test and retest, yet they reported very high reliability, again indicating that the role of examiner is minimal.

The chart-based CSV-1000 is very similar to the Vistech and its latest iteration, the FACT (Vision Science Research Corporation, CA), but there are important differences. As pointed out by Ginsberg, both Vistech and FACT are discrimination tasks where the subject must not only detect the presence of the sinewave grating but also detect its orientation.[28] In contrast, the CSV-1000 uses a detection task where the subject must only detect the presence or absence of a pattern. A number of studies have examined the reliability of the Vistech charts and reported COR values between 0.25 and 0.61, with an average of 0.48, all of which are too low to detect small or subtle differences.[29–33] A few studies have exam-

ined the reliability of the FACT test, which is also reported to be better than that of Vistech, but still low, which is a disappointment given the improvements in the chart design, which include, among others, a much smaller (0.15) step size than the Vistech and a 3-way forced-choice procedure.[8, 29]

Some studies have compared the reliability of CS tests with the well-designed Pelli-Robson (PR) letter contrast chart described earlier.[6, 8–10] Note that the correct detection of a letter is a 26-alternative forced-choice task, and the odds of getting two of three letters in a given triplet correct by chance is extremely low. Thus, subjects are free to guess, but guessing is very unlikely to improve their score. Procedural details, such as the small step size, the psychophysical forced-choice (FC) procedure, and the scoring method, have contrib-

uted to the high reliability of the Pelli-Robson chart that tends to have a COR of about 0.18.[21–26] The ICC has also been reported to be 0.81 or higher, depending on the scoring method used.[27]

In comparison with the PR chart, the step size of the CSV-1000 is comparable but the psychophysical procedure is quite different. The CSV-1000 uses a hybrid technique that combines a 2-alternative forced-choice criterion-free method (top or bottom) with the option to opt out of the FC procedure with a criterion-dependent response (both top and bottom blank). The FC procedure helps to minimize criterion effects between subjects, but with only two alternatives, the subjects can guess the target location correctly 50% of the time for any given contrast level. The probability that subjects will guess correctly twice in a row is 0.25, whereas the probability of being correct three times in a row is about 0.13. Thus, the variability in test–retest results may occur because if a subject is not sure if they detect a grating pattern and guess, then they have a 50% chance of being right. However, on the next administration of the test, the subject may guess wrong. However, if the instructions stress that subjects should immediately indicate when they are no longer sure of the target's location, then guessing will be minimized and test–retest reliability may be improved. We suspect that the reliability differences observed between our study and that of Pomerance and Evans[23] might be due to the instructions given to the observers; they may have stressed to their subjects how important it is to report "no detection" as soon as the two locations appeared homogeneously gray, whereas in our study we instructed the subjects at the beginning of the test that three responses were allowed: top, bottom, or "none."

In support of our conjecture that guessing may be the most important contributor to the low reliability of the test, we have collected pilot data using a modified procedure where subjects were instructed to indicate the target's location regardless of how confident they were that they actually saw a pattern. In other words, subjects were only allowed to report "top or bottom." We tested each spatial frequency twice in this way (one descending trial and then one ascending trial), then scored CS as the lowest contrast correctly detected twice in a row. This simple modification resulted in a significant improvement in the test–retest reliability score, which we are now testing with a larger sample size.

In summary, we report that test–retest reliability of CSV-1000 is low in a sample of young adults and children when following the manufacturer's instructions. We note that our sample size was small and that only visually normal observers were studied. In addition, we did not examine the effect of different examiners on children's COR, but it seems likely that the COR would either be the same or worse than those obtained with the same examiner. The low COR values obtained in this study agree with the generally low COR values reported for other chart-based grating CS tests and highlight the need for either new protocols or new tests that can accurately measure CS in a clinical setting.

ACKNOWLEDGMENTS

REFERENCES

1. Owsley C. Contrast sensitivity. Ophthalmol Clin North Am 2003; 16:171–7.
2. Packer M, Fine IH, Hoffman RS. Contrast sensitivity and measuring cataract outcomes. Ophthalmol Clin North Am 2006;19:521–33.
3. Arditi A. Improving the design of the letter contrast sensitivity test. Invest Ophthalmol Vis Sci 2005;46:2225–9.
4. Pelli DG, Robson JG. Are letters better than gratings? Clin Vis Sci 1991;6:409–11.
5. Pelli DG, Robson JG, Wilkins AJ. The design of a new letter chart for measuring contrast sensitivity. Clin Vis Sci 1988;2:187–99.
6. Haymes SA, Roberts KF, Cruess AF, Nicolela MT, LeBlanc RP, Ramsey MS, Chauhan BC, Artes PH. The letter contrast sensitivity test: clinical evaluation of a new design. Invest Ophthalmol Vis Sci 2006;47:2739–45.
7. Patel PJ, Chen FK, Rubin GS, Tufail A. Intersession repeatability of contrast sensitivity scores in age-related macular degeneration. Invest Ophthalmol Vis Sci 2009;50:2621–5.
8. Buhren J, Terzi E, Bach M, Wesemann W, Kohnen T. Measuring contrast sensitivity under different lighting conditions: comparison of three tests. Optom Vis Sci 2006;83:290–8.
9. Thayaparan K, Crossland MD, Rubin GS. Clinical assessment of two new contrast sensitivity charts. Br J Ophthalmol 2007;91:749–52.
10. Dougherty BE, Flom RE, Bullimore MA. An evaluation of the Mars Letter Contrast Sensitivity Test. Optom Vis Sci 2005;82:970–5.
11. Bodis-Wollner I, Camisa JM. Contrast sensitivity measurement in clinical diagnosis. In: Lessell S, van Dalen JTW, ed. Neuro-ophthalmology. Amsterdam: Excerpta Medica; 1980:373–401.
12. Dos Santos NA, Alencar CC. Early malnutrition diffusely affects children contrast sensitivity to sine-wave gratings of different spatial frequencies. Nutr Neurosci 2010;13:189–94.
13. Li RW, Ngo C, Nguyen J, Levi DM. Video-game play induces plasticity in the visual system of adults with amblyopia. PLoS Biol 2011; 9:e1001135.
14. Caplovitz GP, Kastner S. Carrot sticks or joysticks: video games improve vision. Nat Neurosci 2009;12:527–8.
15. Polat U, Ma-Naim T, Spierer A. Treatment of children with amblyopia by perceptual learning. Vision Res 2009;49:2599–603.
16. Li R, Polat U, Makous W, Bavelier D. Enhancing the contrast sensitivity function through action video game training. Nat Neurosci 2009;12:549–51.
17. Leek MR. Adaptive procedures in psychophysical research. Percept Psychophys 2001;63:1279–92.
18. Hot A, Dul MW, Swanson WH. Development and evaluation of a contrast sensitivity perimetry test for patients with glaucoma. Invest Ophthalmol Vis Sci 2008;49:3049–57.
19. Hou F, Huang CB, Lesmes L, Feng LX, Tao L, Zhou YF, Lu ZL. qCSF in clinical application: efficient characterization and classification of contrast sensitivity functions in amblyopia. Invest Ophthalmol Vis Sci 2010;51:5365–77.
20. Heravian J, Shoeibi N, Azimi A, Yasini S, Ostadi Moghaddam H, Yekta AA, Esmailey H. Evaluation of contrast sensitivity, color vision and visual acuity in patients with and without diabetes. Iran J Ophthalmol 2010;22:33–40.
21. Krasny J, Andel M, Brunnerova R, Cihelkova I, Dominek Z, Lebl J, Papadopoulos K, Soucek P, Treslova L. The contrast sensitivity test in early detection of ocular changes in the relation to the type I diabetes mellitus compensation in children, teenagers, and young adults. Recent Pat Inflamm Allergy Drug Discov 2007;1:232–6.
22. Gandolfi SA, Cimino L, Sangermani C, Ungaro N, Mora P, Tardini MG. Improvement of spatial contrast sensitivity threshold after sur-

gical reduction of intraocular pressure in unilateral high-tension glaucoma. Invest Ophthalmol Vis Sci 2005;46:197–201.

23. Pomerance GN, Evans DW. Test-retest reliability of the CSV-1000 contrast test and its relationship to glaucoma therapy. Invest Ophthalmol Vis Sci 1994;35:3357–61.

24. León A, Estrada J, Quiroz D, Bedoya D. [Reliability of CSV 1000 to evaluate the role of contrast sensitivity in children between seven and ten years.] Cienc Tecnol Salud Vis Ocul 2010;8:1.

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1: 307–10.

26. Owsley C, Sekuler R, Siemsen D. Contrast sensitivity throughout adulthood. Vision Res 1983;23:689–99.

27. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19: 231–40.

28. Ginsburg AP, Cannon MW. Comments on variability in contrast sensitivity methodology. Vision Res 1984;24:287.

29. Pesudovs K, Hazel CA, Doran RM, Elliott DB. The usefulness of Vistech and FACT contrast sensitivity charts for cataract and refractive surgery outcomes research. Br J Ophthalmol 2004;88:11–6.

30. Reeves BC, Wood JM, Hill AR. Vistech VCTS 6500 charts—within- and between-session reliability. Optom Vis Sci 1991;68:728–37.

31. Kennedy RS, Dunlap WP. Assessment of the Vistech contrast sensitivity test for repeated-measures applications. Optom Vis Sci 1990; 67:248–51.

32. Long GM, Tuck JP. Reliabilities of alternate measures of contrast sensitivity functions. Am J Optom Physiol Opt 1988;65:37–48.

33. Rubin GS. Reliability and sensitivity of clinical contrast sensitivity tests. Clin Vis Sci 1988;2:169–77.

Susan A. Kelly
Illinois College of Optometry
3241 S. Michigan Ave
Chicago, IL 60616
e-mail: skelly@ico.edu