O P O 5 8 3 B
Journal Name | Manuscript No.
Dispatch: 7.6.08 | Journal: OPO | CE: Krishna Sarma
Author Received: | No. of pages: 17 | PE: Mahendrakumar

Ophthal. Physiol. Opt. 2008 **28:** 1–17

Technical Note

Repeatability and agreement in the measurement of horizontal fusional vergences

B. Antona, A. Barrio, F. Barra, E. Gonzalez and I. Sanchez

Departamento de Optica II (Optometria y Vision), Universidad Complutense, Madrid, Spain

Abstract

Purpose: This study was designed to determine the repeatability of fusional vergence ranges measured using the rotary prisms in the phoropter and in free space using the prism bar. The level of agreement between the two methods was also investigated..

Methods: In two separate sessions, negative and positive fusional vergence ranges (NFV and PFV, respectively) were measured at distance and near in 61 young adults (mean age 19.74, S.D. 2.5 years) who were unfamiliar with the methods used. Base-in and base-out blur, break and recovery points were sequentially determined. Both sets of measurements were obtained by the same examiner. At each distance, NFV was determined first and then PFV. The repeatability of the tests and agreement between measurements made with the phoropter rotary prisms and the prism bar were estimated by the Bland and Altman method.

Results: For both the phoropter rotary prisms and prism bar, NFV measurements showed better repeatability than PFV at both near and distance. Mean differences recorded for the NFV break and recovery points were non-significant (under 0.5Δ), while those observed for PFV were generally greater than 2Δ. When agreement between the two tests was assessed, it was found that break points were higher when determined using the phoropter rotary prisms, while recovery points were generally higher for the prism bar method. In clinical terms, according to the expected values of the NFV and PFV, agreement between the two techniques can be described as fair, because although mean differences were never greater than 5.5Δ, 95% agreement intervals were as wide as ±8.00Δ for NFV and ±13.19Δ for PFV.

Conclusions: The two methods used to measure fusional vergences showed fairly good inter-session repeatability for measuring NFV but repeatability was reduced for PFV measurements. The level of agreement observed between the two methods was such that their interchangeable use in clinical practice is not recommended.

Keywords: agreement, phoropter rotary prisms, positive and negative fusional vergence, prism bar, repeatability

Introduction

When assessing binocularity, measuring the fusional vergence ranges provides useful information (Daum, 1991; Saladin, 1998). The fusion reflex is responsible for maintaining compensation of a phoria, so knowing what proportion of the total vergence amplitude is needed to compensate a given phoria is of interest to the clinician. Sheedy and Saladin (1977) proved that, in general, the visual system is capable of sustained good performance, as long as no more than two-thirds of the total amplitude is used.

Phoropter rotary prisms and prism bars are commonly used to assess the amplitude of both the positive (convergence) and negative (divergence) fusional vergences. The blur point is a measure of the amount of relative (free from accommodation) fusional vergence. The break measures the total amount of fusional vergence, while the recovery provides information about the person's ability to regain

1 *Received*: xx Xxxxx 200x
Revised form: xx Xxxxx 200x
Accepted: xx Xxxxx 200x

Correspondence and reprint requests to: Beatriz Antona, Escuela Universitaria de Óptica (Universidad Complutense de Madrid), C/ Arcos de Jalón s/n, 28037 Madrid, Spain.
Tel.: +34 91 394 68 94; Fax: +34 91 394 68 85.
E-mail address: bantona@opt.ucm.es

doi: 110.1111/j.1475-1313.2008.00583.x

Table 1. Means and repeatability (COR) of fusional vergence measurements taken on successive sessions with the phoropter rotary prisms and prism bar (n = 61)

Measure	Mean (Δ)	S.D. (Δ)	MD F–I (Δ)	COR (Δ)
Prism bar				
Distance				
NFV brk	8.63	1.94	0.20	±4.00*
NFV rec	6.26	1.82	0.36	±4.00*
PFV blr	12.92	5.22	1.17	±9.73
PFV brk	23.25	7.68	3.85	±13.93
PFV rec	14.50	4.17	2.44	±8.30*
Near				
NFV blr	8.75	3.37	1.54	±6.60*
NFV brk	12.14	3.35	−0.08	±8.05*
NFV rec	9.78	3.02	−0.21	±6.00*
PFV blr	17.08	6.45	2.16	±10.21
PFV brk	28.91	9.09	1.30	±15.00*
PFV rec	19.65	5.98	2.22	±11.50*
Phoropter rotary prisms				
Distance				
NFV brk	9.99	2.36	0.05	±4.82
NFV rec	5.33	2.02	0.47	±3.00*
PFV blr	15.69	6.36	−0.54	±14.00*
PFV brk	24.68	7.35	2.97	±11.29
PFV rec	11.80	5.68	2.63	±12.00*
Near				
NFV blr	11.73	3.80	0.39	±6.43
NFV brk	15.98	4.29	−0.05	±7.00*
NFV rec	8.24	4.04	−0.44	±7.49
PFV blr	22.08	6.83	1.17	±8.83
PFV brk	29.24	8.36	2.08	±7.74
PFV rec	19.22	6.78	4.02	±16.40*

blr, blur; brk, break; rec, recovery; NFV/PFV, negative/positive fusional vergence; MD, mean difference; S.D., standard deviation; COR, coefficient of repeatability; Δ, prism diopter; F, final; I, initial. Values in the COR column marked with an asterisk indicate that this coefficient was replaced with the 95th percentile of the absolute value of the difference.

single binocular vision after diplopia occurs, presumably through reflex fusion with a possible voluntary component (Scheiman and Wick, 2002).

The blur finding indicates that the limit of fusional vergence has been reached and accommodation is no longer held on the target. When positive fusional reserves are measured, the subject normally notices that the fixation target blurs before fusion breaks and diplopia becomes manifest. Upon introduction of the base-out prisms, the eyes are forced to increase their convergence angle. Concurrently, the vergence-driven accommodation increases, with the amount depending in part on the value of the convergence/accommodative convergence (CAC) ratio, while the blur-driven accommodation attempts to reduce its output and thereby maintain the perception of clear (and single) vision. However, at some point, this 'backing off' of accommodation is insufficient and a blurred perception results (Ciuffreda, 1992). When measuring negative fusional reserves at near, the target usually blurs before diplopia because when the eyes are forced to diverge, accommodation relaxes to provide the additional divergence to avoid diplopia. However, when divergence is induced at distance, accommodation is already at its minimum because the patient is emmetropized during measurements and the break point is obtained without blur.

Although there are many reports in the literature on how fusional vergence ranges can be measured using different techniques and stimuli, few studies have examined the repeatability of the different methods available and the equivalence between them. It is generally accepted that when fusional vergence measurements are repeated in an individual, the second measurement can be considerably different than the first (Rouse *et al*., 2002). It has also been demonstrated that fusional ranges are affected by the order in which measurements are taken (Goss, 1995; Rosenfield *et al*., 1995). In the past, several studies have tried to establish the repeatability of vergence ranges in adults and children, especially in terms of the base-out to break end-point but, in some cases, repeatability estimates were based solely on measurements obtained in a single test session and/or the studies also had other purposes, with information on repeatability only obtained indirectly or as a secondary goal (Feldman *et al*., 1989; Penisten *et al*., 2001; Scheiman *et al*., 2005).

In this study, we examined maximum horizontal fusional vergence amplitudes measured in healthy adults using the phoropter rotary prisms and a prism bar. Rotary prisms are ideal for smoothly modifying the prism demand and provide fairly repeatable results in young adults (Penisten *et al*., 2001). It has also been noted, however, that results are less reliable in children and that inter-examiner variation is high (Rouse *et al*., 2002). Measuring vergence ranges in free space using a prism bar more closely resembles habitual conditions. This method is especially useful for measurements in young children because eye movements can be seen so that the clinician can objectively confirm the child's replies.

The aim of this study was to determine the degree of repeatability of these measurements when using the phoropter rotary prisms and the prism bar. Agreement between measurements made using the two methods was also investigated.

Methods

Study population

The study population was comprised of 61 subjects aged 18–32 years (mean 19.74, S.D. 2.5 years) recruited randomly by informative talks from the first year students of the School of Optics, Universidad Complu-

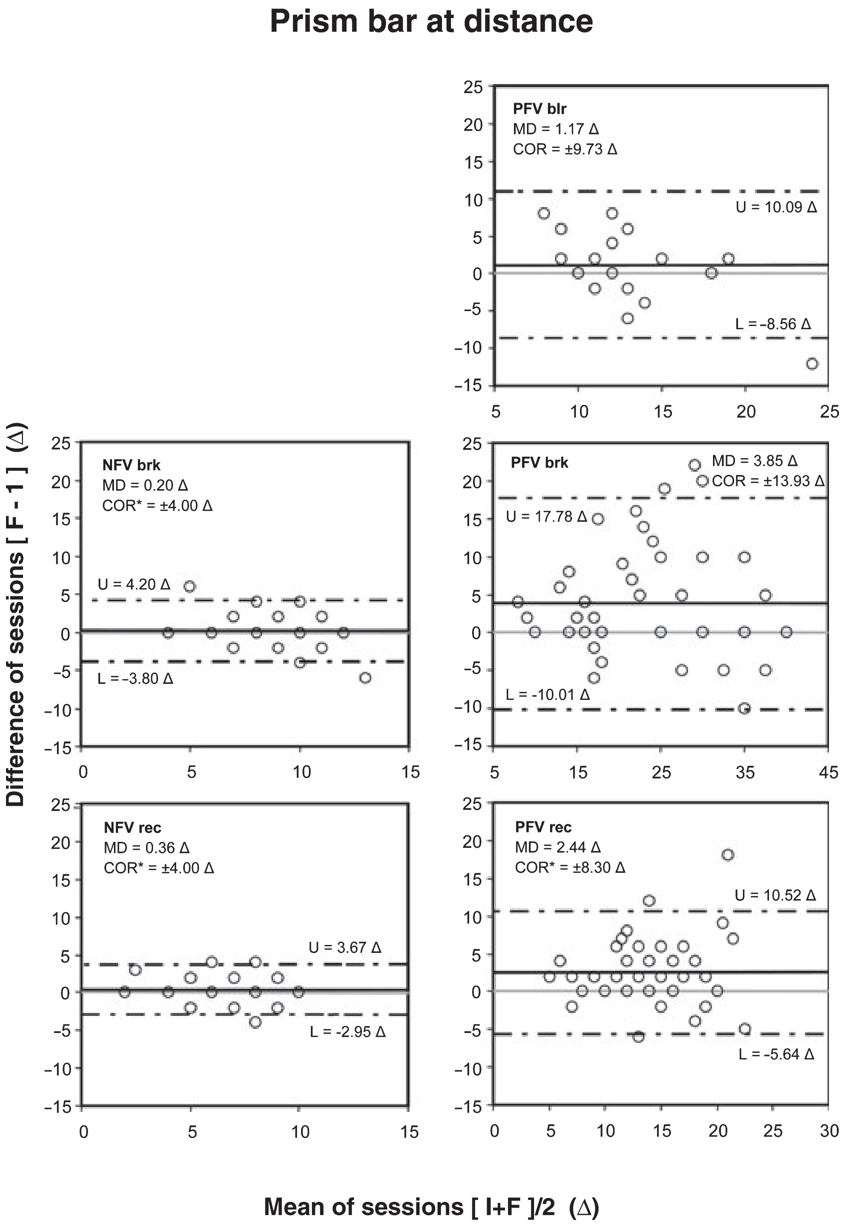

Figure 1. Difference vs mean plots of repeatability for prism bar measurements of NFV (left) and PFV (right) at distance. The solid line represents the averaged difference of the measurements from final session and initial session (mean difference, MD). The dotted lines indicate the lower (L) and the upper (U) 95% limits of agreement (MD ± 1.96 S.D.). When the distribution of differences was not normal, the 95th percentile of the absolute differences (COR*) was used.

tense de Madrid, Madrid, Spain. The subjects were selected because they were unaccustomed to the type of tests performed, being recently admitted students. The results of this study could therefore be extrapolated to a random clinical population of this age range and with similar near work demand. The study design fulfilled the tenets of the Declaration of Helsinki. The clinical criteria for inclusion were:

(1) A corrected visual acuity (VA) greater or equal to 0.9 decimal visual acuity of Snellen (20/22) in each eye at distance and near.
(2) No ocular pathology.
(3) No history of refractive surgery, strabismus, nystagmus or amblyopia.
(4) No medication or disease that could affect accommodation, fusional vergences or ocular motility.
(5) Asymptomatic with no accommodative or vergence anomalies. The criteria used to diagnose these dysfunctions were those used in the integrative analysis approach by Scheiman and Wick (2002). Binocular test results were compared with the normal value, and we considered that subjects were without accommodative or vergence anomalies if

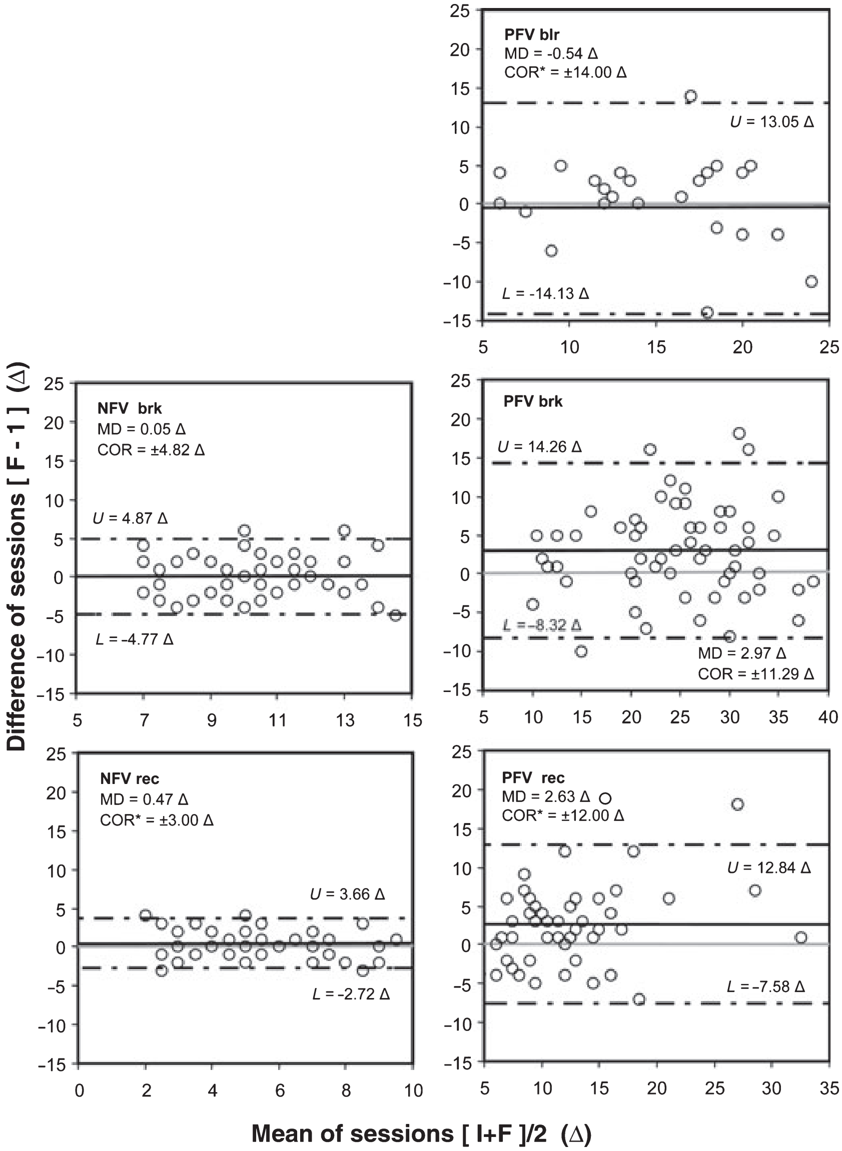

Figure 2. Difference vs mean plots of repeatability for phoropter rotary prisms measurements of NFV (left) and PFV (right) at distance. The solid line represents the averaged difference of the measurements from final session and initial session (mean difference, MD). The dotted lines indicate the lower (L) and the upper (U) 95% limits of agreement (MD ± 1.96 S.D.). When the distribution of differences was not normal, the 95th percentile of the absolute differences (COR*) was used.

they were free from any binocular vision symptoms and there were no findings diagnostic of any binocular anomaly.

Test procedures

Following history and symptoms, some optometric characteristics of each subject were determined in the following tests:

(1) Monocular and binocular VA with and without correction. We used Snellen optotypes projected at distance (6 m) and printed at near (40 cm). Habitual correction was also recorded.
(2) Keratometry and objective refraction were determined using a Topcon KR 7000P autorefractometer–keratometer (Topcon Corporation, Tokyo, Japan).
(3) Subjective refraction was determined using the normal procedure with Snellen optotypes projected at 6 m and a manual phoropter. Subjective refraction was performed by means of a monocular fogging method with cross-cylinder, followed by binocular balancing to a standard end-point of maximum plus for best visual acuity (Carlson and Kurtz, 2004).

Of the initial 77 subjects screened, 64 subjects who fulfilled the first four inclusion criteria underwent the

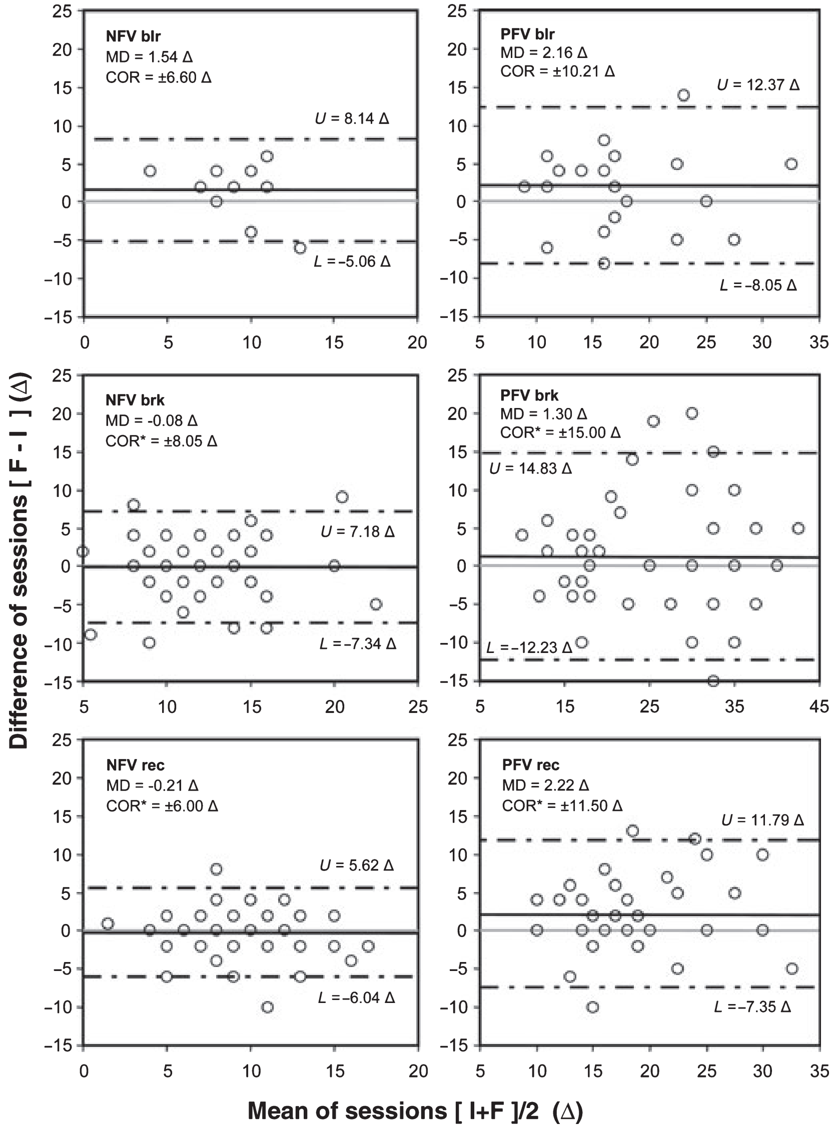

Figure 3. Difference vs mean plots of repeatability for prism bar measurements of NFV (left) and PFV (right) at near. The solid line represents the averaged difference of the measurements from final session and initial session (mean difference, MD). The dotted lines indicate the lower (L) and the upper (U) 95% limits of agreement (MD ± 1.96 S.D.). When the distribution of differences was not normal, the 95th percentile of the absolute differences (COR*) is used.

first session of measurements, in whom horizontal fusional vergences as well as the additional binocular vision tests were conducted. The tests performed were: horizontal phorias far and near using the von Graefe technique; monocular amplitude of accommodation with minus lenses, monocular-estimate-method dynamic retinoscopy, binocular accommodative facility with ±2.00 D flipper lenses, negative and positive relative accommodation, and stereoacuity using the Randot and the TNO tests. All these measurements were made with the subjective refractive correction in place. A different clinician considered the test results of the initial session of each participant to rule out subjects with accommodative or vergence anomalies.

Of the 77 subjects screened, 64 passed to the first testing session. We identified one subject with accommodative insufficiency and another one with convergence insufficiency who where excluded from the study. All subjects presented for the second session of measurements but we excluded one subject who started orthokeratology treatment between the initial and the final session.

According to Bland and Altman (1986), the best way to assess the repeatability of an instrument is to take several measurements in a series of subjects. Thus, negative fusional vergence (NFV) and positive fusional vergence (PFV) ranges were measured at distance and near in two sessions separated by a time interval of at

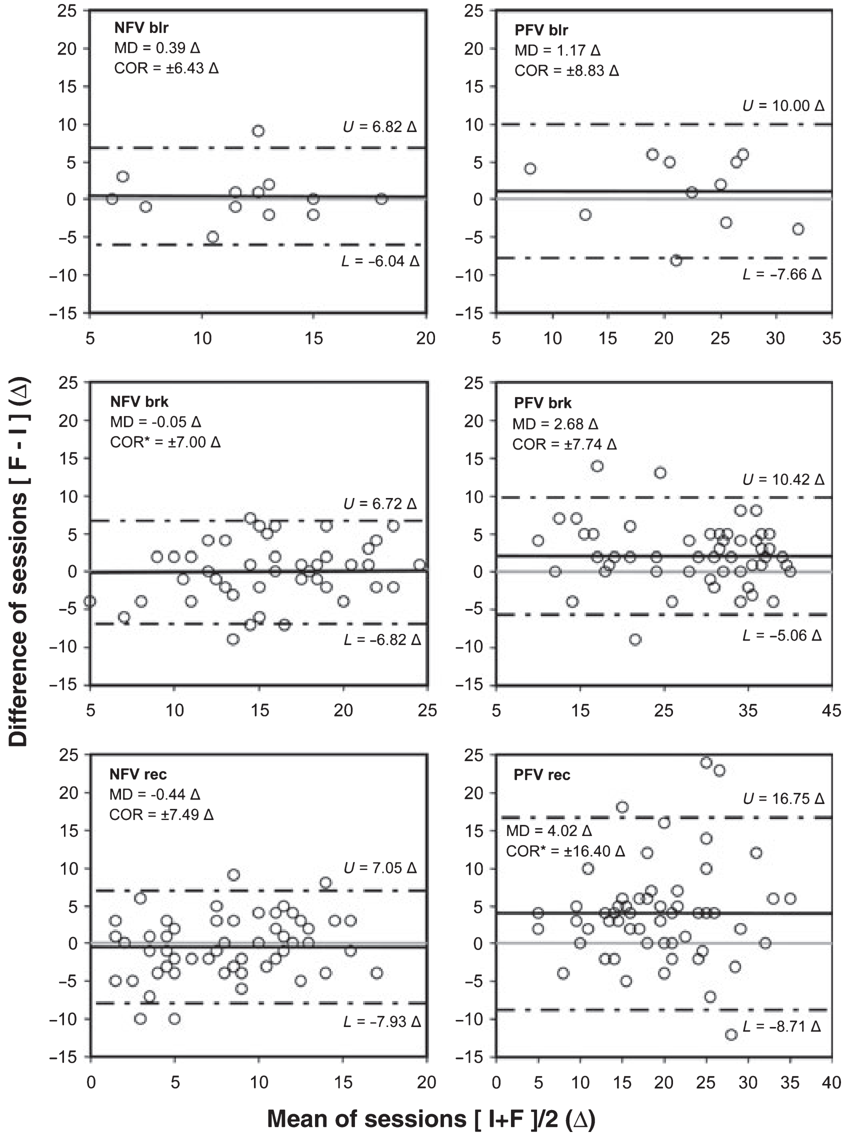

Figure 4. Difference vs mean plots of repeatability for phoropter rotary prisms measurements of NFV (left) and PFV (right) at near. The solid line represents the averaged difference of the measurements from final session and initial session (mean difference, MD). The dotted lines indicate the lower (L) and the upper (U) 95% limits of agreement (MD ± 1.96 S.D.). When the distribution of differences was not normal, the 95th percentile of the absolute differences (COR*) is used.

least 24 h with a maximum of 10 days between sessions. We chose this time interval between sessions because it was short enough for there to be a small probability that the subject suffered real changes in their visual abilities, but long enough to avoid the possibility of a significant learning effect (Argimon and Jimenez, 2004).

In the first session, written consent to participate was obtained from each subject after informing them of the aims of the study. The tests in the two sessions were undertaken by the same examiner. The results of the first set of measurements were not available during the second session, to avoid any possible examiner bias. Vergence ranges were first measured at distance and then at near. For each of the two distances, NFV was determined before PFV and we made only one measurement of NFV and PFV during each session. The fixation target was a column of letters isolated from the Snellen chart corresponding to a decimal visual acuity of 0.8. For measurements at distance, the target was projected at 6 m and for near, the target was presented as a card placed at 40 cm. Room lighting was used for measurements at distance and a light source directed towards the card was added for the tests at near, making sure that no shadows were produced. Subjects were allowed to rest for 15 s between each measurement, during which time they were instructed to gaze into the distance (6 m) to minimize prism adaptation effects (Schor and

Table 2. Agreement between fusional vergence measurements made with the phoropter rotary prisms and prism bar (n = 61)

	Distance		Near	
Measure	**MD ph−pb(Δ)**	**COA (Δ)**	**MD ph−pb(Δ)**	**COA (Δ)**
NFV blr	–	–	3.66	±8.00*
NFV brk	1.24	±3.54	3.84	±6.81
NFV rec	−0.99	±3.71	−1.54	±7.02
PFV blr	2.99	±9.48	5.05	±13.19
PFV brk	1.39	±10.06	0.33	±11.18
PFV rec	−2.61	±10.50*	−1.88	±9.00*

blr, blur; brk, break; rec, recovery; NFV/PFV, negative/positive fusional vergence; MD, mean difference; COA, coefficient of agreement; Δ, prism diopter; ph, phoropter rotary prisms; pb, prism bar. Values in the COA column marked with an asterisk indicate that this coefficient was replaced with the 95th percentile of the absolute value of the difference.

Ciuffreda, 1983). Each subject was requested to pick out a numbered ball to indicate the order of the tests (phoropter rotary prisms, prism bar), which was randomized to avoid the learning effect and/or subject fatigue influencing the results. Each test was performed in exactly the same manner in each subject and at approximately the same time of day. The second set of tests was similarly scheduled.

During the tests, the subject was instructed to make a effort to keep the target letters as focused as possible and to inform the examiner the moment they noted the letters became constantly blurred (blur), the moment the column of letters became double and the image could not be reunited (break) and then when a single column was again observed (recovery). To help the subject identify the blur point, before starting the measurement procedure, blur was simulated by adding +0.50 D to the subject's distance correction. Further, given that the criterion used for blur was the first sustained blur noted, when the subject manifested the start of blur, the examiner momentarily stopped increasing prism power and the subject was asked to confirm that the target could not effectively be cleared. The break point was taken as the lowest prism power at which the subject reported sustained double vision, that is, the maximum vergence amplitude was sought. Once the image had doubled, the dioptric value was increased by 4 or 5Δ, and then reduced at the same rate until the target became single again (recovery value).

Fusional vergences using the rotary prisms of the phoropter. With the subject's distance prescription in place in the phoropter, inter-pupilary distance was adjusted. The rotary prisms were placed before each eye and set up for horizontal power ('0' at 90 degrees). Starting from zero, prism power before each eye was simultaneously increased. The amplitude is measured through a smooth gradual increase in prism power rather than discrete increases, as with the prism bar.

Fusional vergences using the prism bar. The subject's distance prescription was placed in a trial frame. During measurements, a standard horizontal prism bar was used which had the following prism vergence steps: 1, 2, 4–20 in 2Δ steps and 25–40 in 5Δ steps. It was placed in front of the subject's right eye in the spectacle plane and prism strength was increased at a speed of one step every 2 or 3 s. It is relatively easy to induce undesired vertical prism through an unintended tilt of the prism bar, especially at higher values, so careful monitoring of orientation was essential. Although it is expected that a person will be capable of noticing the blur point when

Table 3. Expected values for fusional vergences taken with phoropter rotary prisms (Morgan, 1944) and with prism bar (Wesson, 1982)

	Phoropter rotary prism		Prism bar	
	This study n = 61	**(Morgan, 1944) n = 800**	**This study n = 61**	**(Wesson, 1982) n = 116**
Measure	**Mean (S.D.) (Δ)**			
Distance				
NFV brk	9.99 (2.36)	7 (3)	8.63 (1.94)	7(3)
NFV rec	5.33 (2.02)	4 (2)	6.26 (1.82)	4 (2)
PFV blr	15.69 (6.36)	9 (4)	12.92 (5.22)	–
PFV brk	24.68 (7.35)	19 (8)	23.25 (7.68)	11 (7)
PFV rec	11.80 (5.68)	10 (4)	14.50(4.17)	7 (2)
Near				
NFV blr	11.73 (3.80)	13 (4)	8.75 (3.37)	–
NFV brk	15.98 (4.29)	21 (4)	12.14 (3.35)	13 (6)
NFV rec	8.24 (4.04)	13 (5)	9.78 (3.02)	10 (5)
PFV blr	22.08 (6.83)	17 (5)	17.08 (6.45)	–
PFV brk	29.24 (8.36)	21 (6)	28.91 (9.09)	19 (9)
PFV rec	19.22 (6.78)	11 (7)	19.65 (5.98)	14 (7)

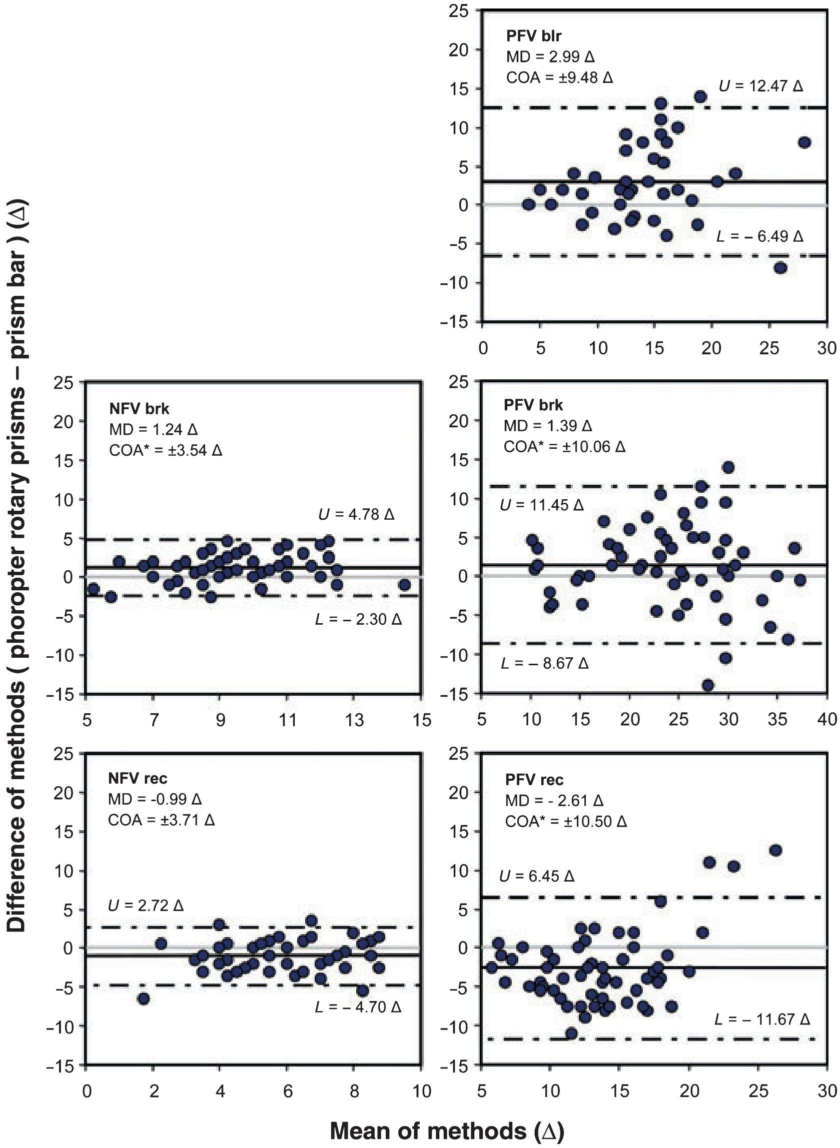

Figure 5. Difference vs mean plots of method comparisons for distance vision (6 m). The solid line represents the averaged difference (MD) of the measurements from phoropter rotary prisms and prism bar. The dotted lines indicate the lower (L) and the upper (U) 95% limits of agreement (MD ± 1.96 S.D.). When the distribution of differences was not normal, the 95th percentile of the absolute differences (COA*) is used.

measuring PFV at near and distance, and NFV at near, in practice, many subjects find it difficult to recognize blurring of the test image because the increasing vergence steps correspond to several prism diopters.

Statistical analyses

Data analysis was performed using the Analyse-it for Microsoft Excel program (Analyse-it Software Ltd, Leeds, UK http://www.analyse-it.com).

The Bland and Altman method was used to determine both the repeatability of the tests and agreement between them (Bland and Altman, 1986; Zadnik *et al.*, 1992). In diagnostic terms, the advantage of this method is that test agreement is expressed in the same units of measurement as the test itself and allows the clinician to establish his own criterion as to whether or not a difference is significant. This method was used when the differences, as established by the Anderson–Darling normality test (D'Augostino and Stevens, 1986), showed a normal distribution.

The factors determined were the mean difference, the standard deviation (S.D.), the coefficient of repeatability (COR = 1.96 × S.D.) and the limits of agreement at the 95% level (mean difference ± COR). The level of significance was set at $p < 0.05$. Although with a relatively large sample of subjects the impact of a non-normal distribution over the results is expected to be small, but if a non-normal distribution was found, we determined the 95th

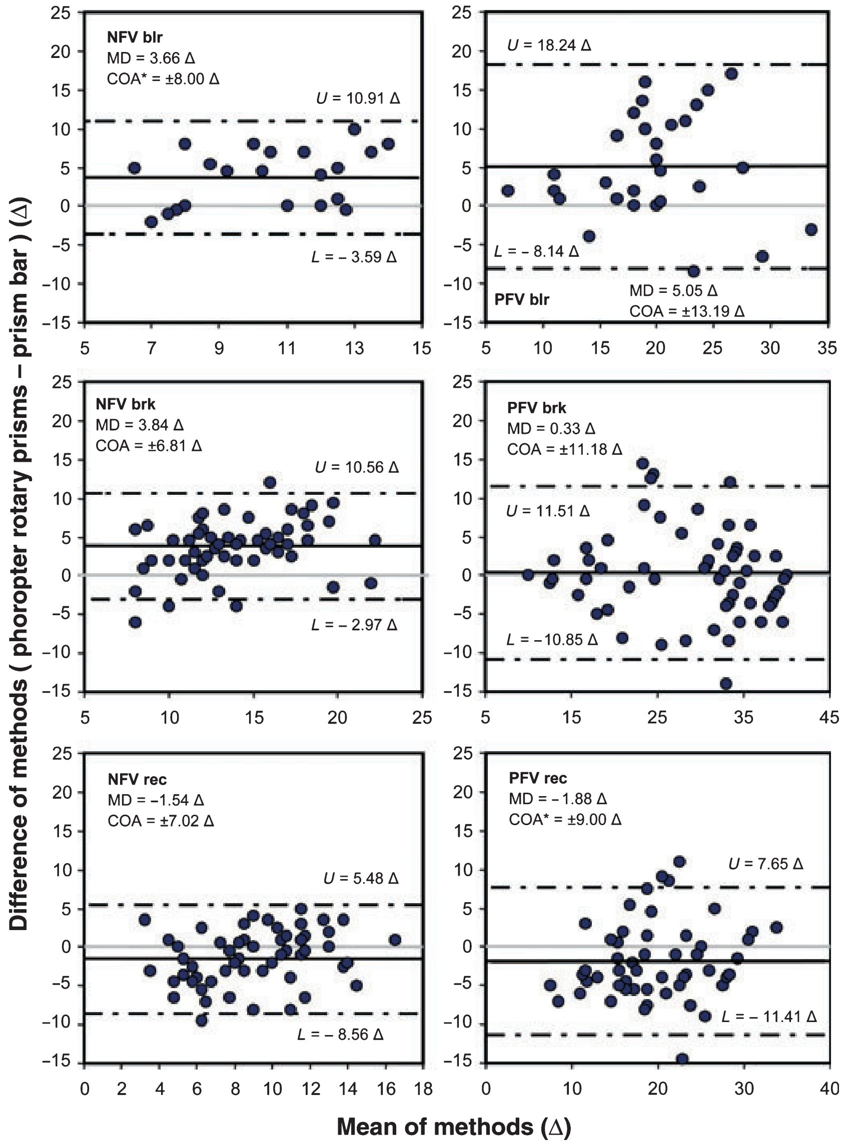

Figure 6. Difference vs mean plots of method comparisons for near vision (40 cm). The solid line represents the averaged difference (MD) of the measurements from phoropter rotary prisms and prism bar. The dotted lines indicate the lower (L) and the upper (U) 95% limits of agreement (MD ± 1.96 S.D.). When the distribution of differences was not normal, the 95th percentile of the absolute differences (COA*) is used.

percentile of the absolute values of the differences instead of calculating the COR. Similarly, we determined the coefficient of agreement (COA) between the phoropter rotary prisms and prism bar measurements. The limits of the agreement interval constitute a threshold for the differences in successive measures that have to be surpassed if the difference indicates that a change in the value has in effect occurred and cannot simply be explained by natural variation among measurements.

Given the sample size, a small difference may be statistically significant yet not clinically significant. Differences from the mean were plotted to establish the limits of agreement at the 95% level and obtain a better idea of the repeatability of the measures.

Results

The range of ametropia of the 61 subjects who participated in both testing sessions was from −0.50 to −6.50 D of myopia and from +0.50 to +1.00 D of hyperopia and up to –2.75 D of astigmatism. The number of subjects in each heterophoria categories was: exophores: 22 for 6 m and 29 for 40 cm; orthophores: 11 (6 m and 40 cm); esophores: 28 for 6 m and 21 for 40 cm.

Repeatability

Mean time between sessions was 7.01 days (S.D. = 1.60 days). Using the phoropter rotary prisms

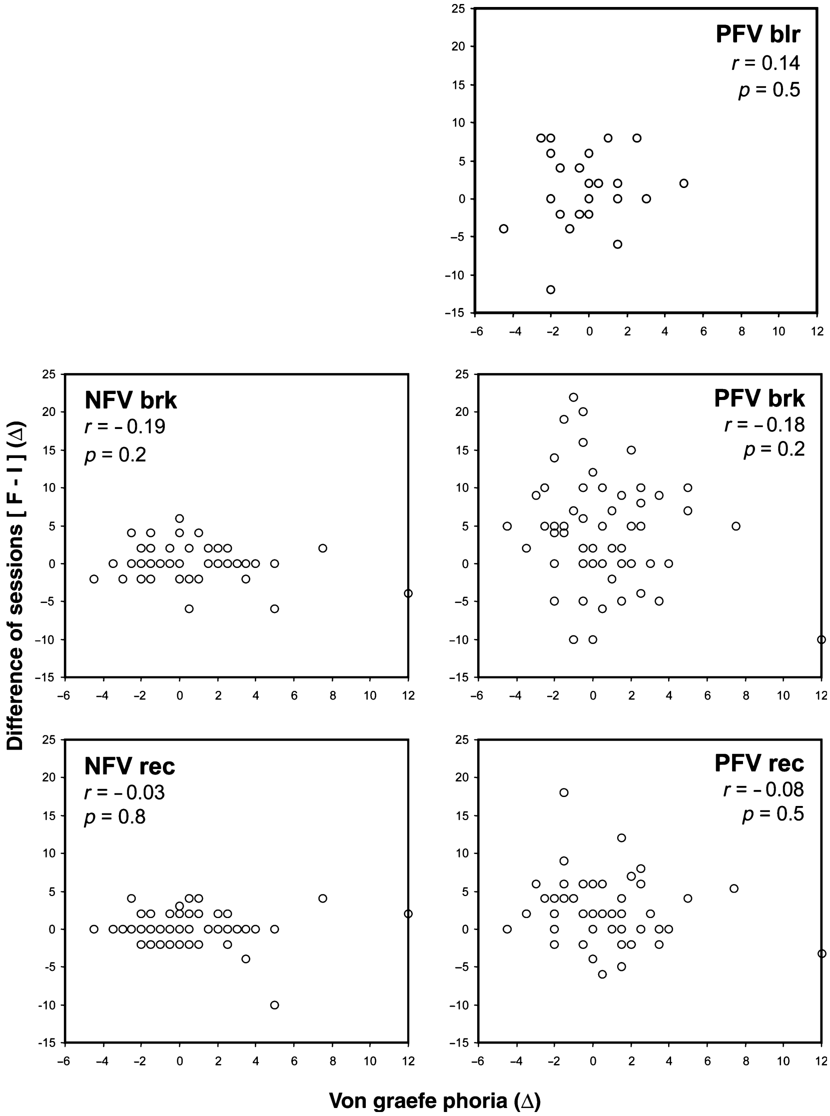

Figure 7. Difference between sessions (final–initial) vs phoria plots for prism bar measurements of NFV (left) and PFV (right) at distance. The Pearson's coefficient of correlation (*r*) is shown.

and prism bar, mean measurement differences and 95% concordance intervals were lower, that is, repeatability was better for base-in (divergence) fusional reserve measurements (NFV) than for base-out (convergence) ranges (PFV) both at near and far.

Table 1 shows the summary of repeatability measures for phoropter rotary prisms and prism bar. Mean inter-session differences recorded for NFV break and recovery points were so small that they may be considered clinically non-significant (under 0.5Δ), while differences between PFV ranges measured in the initial and final sessions were generally greater than 2Δ. The results are shown for prism bar in *Figure 1* (for 6 m) and in *Figure 3* (for 40 cm) and for phoropter rotary prisms in *Figure 2* (for 6 m) and in *Figure 4* (for 40 cm).

Figures 1–4 are difference vs mean plots with the difference between two measures (final–initial) on the y-axis plotted against the average of the two measures on the x-axis. For each plot, if measures show good repeatability, the averaged difference will be close to 0, and the ± 1.96 S.D., or the 95% limits of agreement will be small. None of the plots shows a tendency for the difference to increase with the prismatic value, i.e. the repeatability of the tests does not change with size of the fusional range.

Agreement between methods

When the degree of agreement between the two procedures was examined, it was noted that blur and break points determined using the phoropter rotary prisms

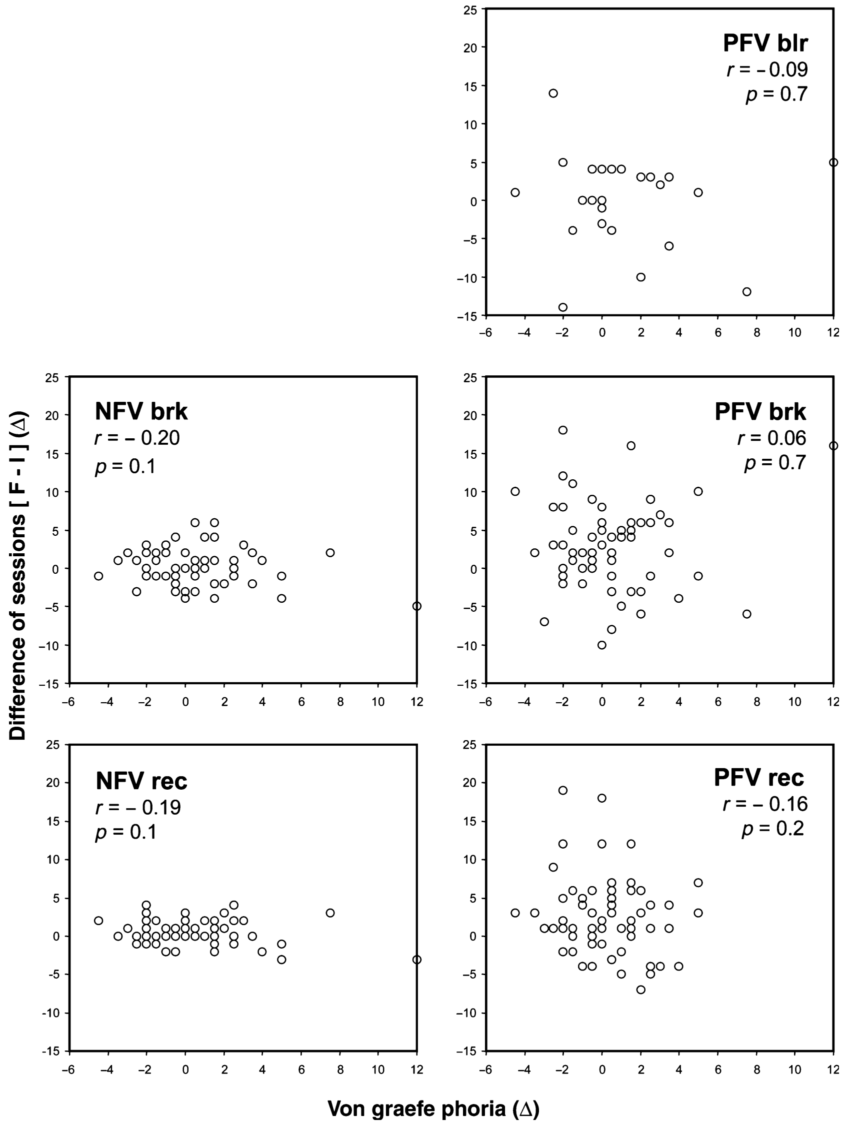

Figure 8. Difference between sessions (final−initial) vs phoria plots for phoropter rotary prisms measurements of NFV (left) and PFV (right) at distance. The Pearson's coefficient of correlation (*r*) is shown.

were higher, while recovery values were higher when measured using the prism bar (see *Table 2*). These findings indicate that the target appears blurred and fusion is broken with a smaller prism when the prism bar is used rather than the phoropter rotary prisms, although fusion is recovered at a higher prism value with the prism bar.

Agreement between the two techniques can be described as fair in clinical terms, because although mean differences did not exceed 5.5Δ, agreement coefficients were high, compared with the expected values of the NFV and PFV (Morgan, 1944; Wesson, 1982) (see *Table 3*). The lowest mean difference was obtained for PFV break at near (0.3Δ). Notwithstanding, it was precisely for this variable that the 95% agreement interval (±11.2Δ) was among the most exaggerated. Blur points also exhibited high mean differences and COA. In effect, the highest mean difference was observed for PFV blur values measured at near, for which measurements made with the phoropter rotary prisms were notably higher than those provided by the prism bar [mean of the differences (MD) = 5.1Δ]. It should be considered, however, that not all participants reported this blurring of the image before reaching the break point ($n = 40$ for PFV at distance, $n = 22$ for NFV at near and $n = 29$ for PFV at near), which could compromise the statistical power of the blur results.

Bland and Altman (1986) suggested that rather than determining correlation coefficients, a more informative method of comparing the results of two methods of testing is to plot the differences between the results

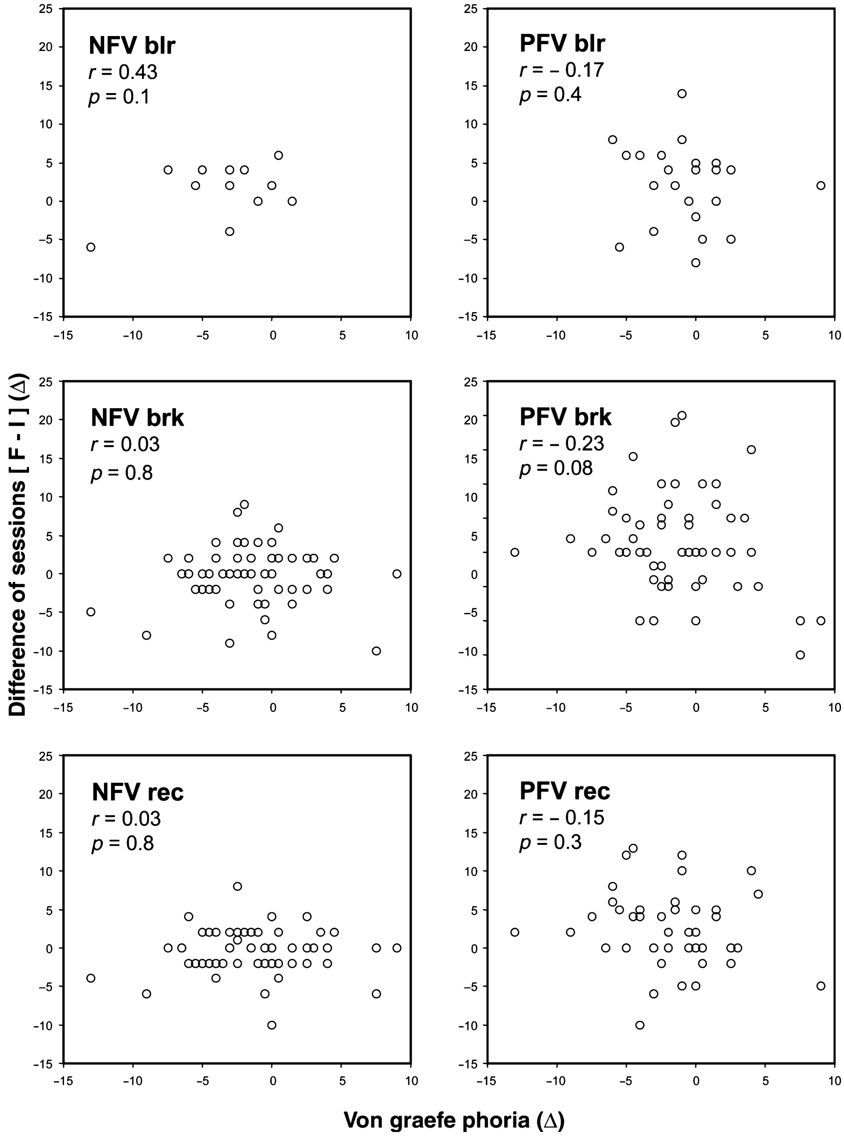

Figure 9. Difference between sessions (final–initial) vs phoria plots for prism bar measurements of NFV (left) and PFV (right) at near. The Pearson's coefficient of correlation (r) is shown.

obtained by the two methods against the mean of the two methods. When this is done, the mean of the results are plotted along the x-axis and the differences are plotted along the y-axis. Horizontal lines are plotted, indicating the MD and the limits of agreement between the two methods, MD ± 1.96 S.D. of the differences. Such plots for our data, shown in *Figure 5* (for 6 m) and *Figure 6* (for 40 cm), indicate high COA, especially for PFV measurements. These COA are sufficiently high so that the two methods could not be used interchangeably. The level of agreement between the tests does not change with the negative fusional range (i.e. the difference does not increase with a higher mean value), but there is some tendency to higher COAs when positive fusional range increased.

Correlation between difference and phoria

Figures 7–10 show the difference between the two measures (final−initial) on the y-axis plotted against the phoria of the subject on the x-axis. *Figures 11 and 12* show the difference between the two methods (phoropter rotary prisms−prism bar) on the y-axis plotted, against the phoria of the subject on the x-axis. None of the plots shows a tendency for the difference to increase with the phoria value, i.e. the repeatability and the agreement between tests do not change with the size of the phoria. The Pearson's correlation coefficient was not significant in any case ($r \le \pm 0.43$, $p > 0.05$).

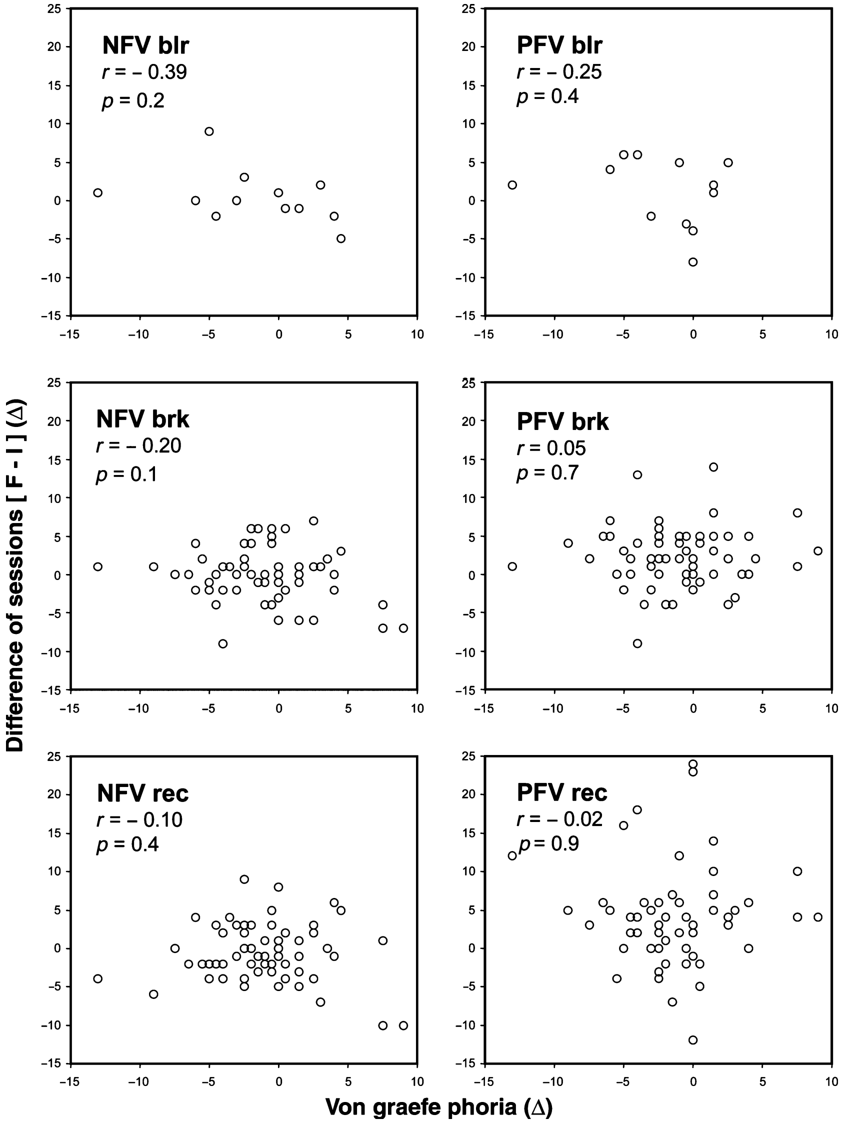

Figure 10. Difference between sessions (final–initial) vs phoria plots for phoropter rotary prisms measurements of NFV (left) and PFV (right) at near. The Pearson's coefficient of correlation (r) is shown.

Discussion

Comparing the mean fusional vergence measurements obtaining in our population with the expected values from Morgan (1944) and Wesson (1982), our subjects appeared to have greater PFV mean values. However, the NFV mean values were quite similar or even smaller than the expected values (see *Table 3*).

Repeatability

Studies examining the repeatability of vergence testing in multiple sessions (Feldman *et al.*, 1989; Penisten *et al.*, 2001; Rouse *et al.*, 2002; Ciuffreda *et al.*, 2006) have yielded conflicting results. Such discrepancies can have clinical implications because only when the expected variability of the measurement method used is understood, it will be possible to recognize true gains made in response to some form of treatment. For example, in two sessions 1 week apart, Rouse *et al.* (2002) tested over 20 children aged 10–11 years and found that an intra-examiner difference of 12Δ would be needed to indicate a valid treatment-based improvement in PFV vergence ranges measured with phoropter rotary prisms at near. With a similar design and data analysis, this study has found similar repeatability results for PFV recovery to those of Rouse *et al.* (2002). However, they found higher PFV break COR results (±14.07Δ for examiner 1 and ±12.00Δ for examiner 2) than in our study (COR = ±7.74Δ). Although more research is needed to confirm if children's responses are less repeatable than those of adults, one might expect that

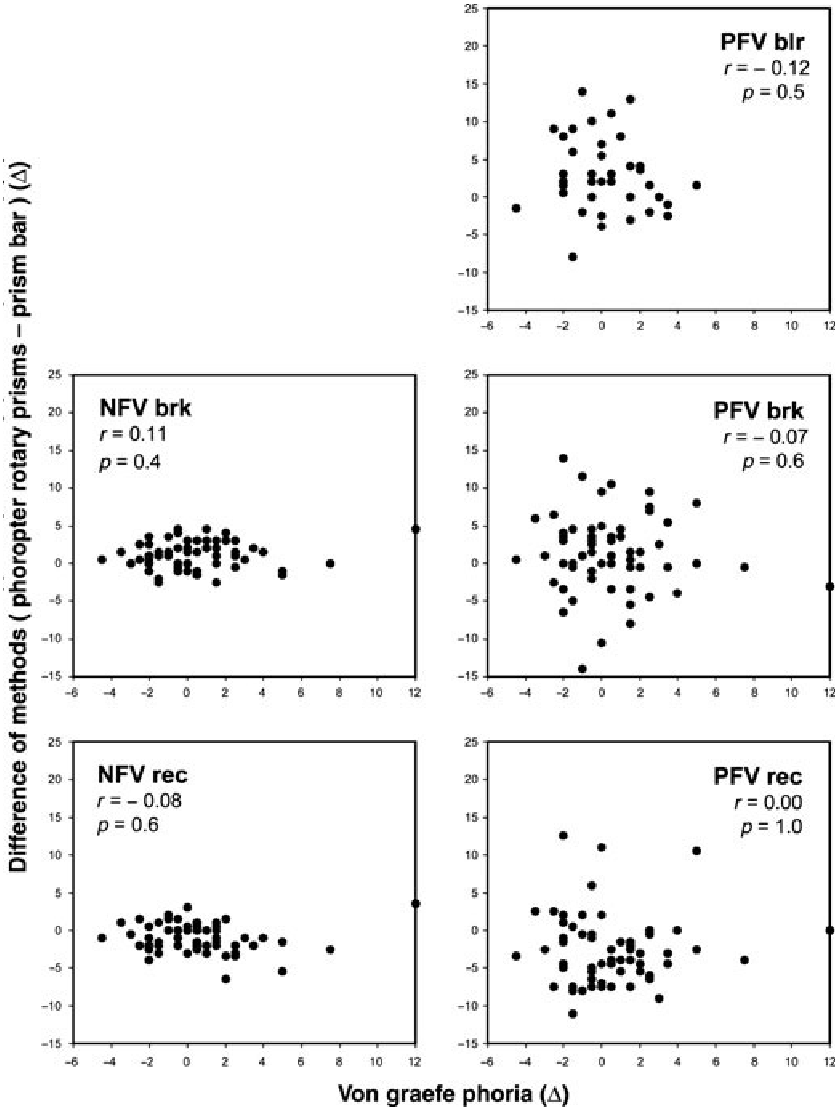

Figure 11. Difference between methods (phoropter rotary prisms–prism bar) vs phoria plots for distance vision (6 m). The Pearson's coefficient of correlation (*r*) is shown.

children might have more difficulty than adults with the psychophysical aspects of this test (poorer observers, more trouble understanding instructions, less reliable subjective end-point reports).

Sheedy and Saladin (1983) claim that it is not uncommon to find a difference of 10Δ between one vergence amplitude measurement and another, unless strict control measures are applied. Among the factors that may affect the repeatability of vergence range measurements are accommodation and proximal convergence. The fact that these two variables are kept at their minimum during measurements at distance could account for the lower variability of the tests when measurements are made at distance.

The repeatability results observed here for the phoropter rotary prisms method of determining fusional vergences (see *Table 1*) indicate less intra-subject variability for NFV measurements compared with PFV. For instance, our smallest COR corresponded to NFV break and recovery at distance (±3Δ and ±4.8Δ, respectively), and the largest COR (±16.4Δ) was obtained for PFV recovery at near, and followed by the intervals recorded for PFV blur recovery and break at distance (±14Δ, ±12.7Δ and ±11Δ, respectively). The smaller agreement intervals observed here for the NFV measurements compared with PFV could be explained by the fact that NFV and PFV functions have somewhat different underlying neural mechanisms. For example, a brain lesion can produce paralysis of one function but not the other (Tannen and Ciuffreda, 1995; Ciuffreda *et al.*, 2006). It is possible that the binocular status of the subjects affects the repeatability of the measure, but the number of esophores and exophores was very similar in our study, so it does not seem to

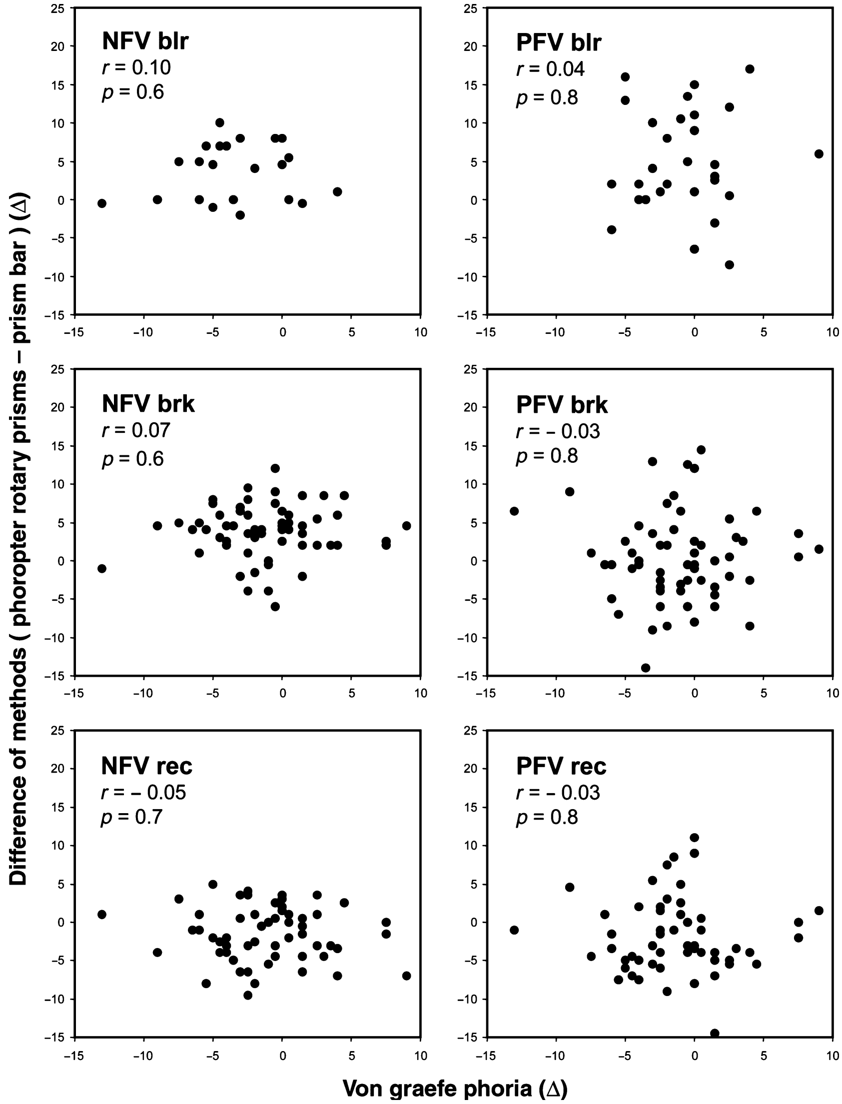

Figure 12. Difference between methods (phoropter rotary prisms–prism bar) vs phoria plots for near vision (40 cm). The Pearson's coefficient of correlation (*r*) is shown.

influence the variability of PFV vs NFV measures. We tested the hypothesis that a greater degree of difference is related to a greater degree of phoria (*Figure 7–12*). This correlation study revealed that the difference between the two measures (final−initial) or between the two methods (phoropter rotary prisms−prism bar) was not greater for higher values of phoria (esophores or exophores), i.e. difference and phoria did not correlate.

Other repeatability studies were carried out with small adult samples and they did not use the Bland and Altman approach, which makes direct comparison difficult. For example, Penisten *et al.* (2001) reported on a sample of eight subjects with a mean age of 28.9 years (S.D. 7.0) who underwent four testing sessions, at which repeated measurements of fusional vergence ranges with phoropter rotary prisms were taken. The average S.D. for each fusional vergence test results (NFV and PFV; blur, break and recovery; distance and near vision) was calculated. NFV average S.D. was between 1.1 and 2.3Δ, and PFV average S.D. was between 1.7 and 2.8Δ. Penisten *et al.* (2001) reported similar qualitative results to this study, in that the variability for each fusional vergence test results was always higher for the PFV measurements with the only exception being the break point at near.

More recently, Ciuffreda *et al.* (2006), undertook a weekly assessment of the repeatability of near vergence ranges determined either in free space (prism bar) or with the phoropter (rotary prisms) over a 10-week period. Three experienced adult subjects (the authors) aged 26, 31 and 57 years were tested and both intra-session and inter-session variability were examined. Obviously, the test conditions in which each study was performed should be considered. Subjects were attended for a total of 10 sessions and three measurements were taken in each session. For every subject, mean and S.D. was calculated without considering different sessions. Then, the agreement interval for each method was calculated from the mean of all the individual S.D. We would expect this method to filter out extreme measures. In addition, their small sample size of three subjects will diminish the statistical power of the results. Thus, the authors acknowledge that the results of their study can probably be interpreted as the best-case scenario and even so, the variability in their results (average S.D.: prism bar: 1.5–5.7Δ; rotary prisms: 2.3–3.9Δ) was similar to the study of Penisten *et al*. (2001) and not too far from the results of this study (S.D.: 1.8–9.1Δ, see *Table 1*).

Agreement between methods

Smooth (rotary prisms) and step (prism bar) vergence testing are both designed to evaluate fusional vergence amplitude and yet several studies (Wesson, 1982; Scheiman *et al*., 1989) have demonstrated that findings are different for smooth vs step vergence.

In the study by Ciuffreda *et al*. (2006) in which vergences were measured in three experienced subjects, all PFV break and recovery averages for the prism bar were higher (break: 39.1Δ; recovery: 38.0Δ) than for the phoropter rotary prisms (break: 32.3Δ; recovery: 29.3Δ). However, NFV break and recovery averages were similar for both the free space and phoropter test methods. The authors proposed that the higher vergence ranges obtained in free space could be due to the influence of peripheral fusion on fusion vergence ability (Burian, 1939; Tannen and Ciuffreda, 1995). It is interesting to see that here both for PFV and NFV measurements, mean blur and break values were higher for the phoropter rotary prisms method, while recoveries were higher when measured with the prism bar. This behaviour could be attributed to a greater ease in achieving higher blur and break point values when prism strength is gradually increased and introduced binocularly, as occurs with the rotary prisms of the phoropter. In contrast, the prism bar test requires asymmetrical vergence and step vergence-type changes in prism demand (the prism is introduced monocularly), which may be more difficult for the subject. It is also likely that the effect of peripheral vision is more influential when trying to recover single vision following diplopia, which could explain the higher recovery point found with the prism bar compared with the phoropter rotary prisms. On the other hand, our results have shown that there is some tendency to higher COAs when positive fusional range increased (see *Figures 5 and 6*), probably because the steps between the prism demands have become larger (e.g. 5Δ) compared with the small 1Δ changes in the rotary prisms at these higher values.

Because of the homogeneous clinical characteristics of the sample, the results of this study could be only directly extrapolated to a random clinical population of this age range and with similar near work demand. We consider that more research is needed in this area to confirm our results for a general population and research to determine the influence of inter-examiner repeatability and to determine if children's responses are less repeatable than those of adults.

Conclusions

The two methods of measuring fusional vergence assessed here can be described as fairly reliable for determining NFV, although PFV measurements show low repeatability. In clinical terms, the variability in vergences observed for each procedure should be taken into account when determining binocular vision status or assessing treatment progress.

In clinical terms, the fair agreement observed between the two methods according to the expected values of the NFV and PFV means that the clinician should be cautious when comparing prism bar and rotary prisms vergence measurements, especially base-out vergence ranges. These findings are not unexpected and the clinician in practice knows that it is important that the method selected for the initial evaluation is also used for all future measurements.

References

Argimon, J. M. and Jimenez, J. (2004) *Métodos de Investigación Clínica y Epidemiológica*. Elsevier, Madrid.

Bland, J. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1,** 307–310. 5

Burian, H. M. (1939) Fusional eye movements: role of peripheral retinal stimuli. *Arch. Ophthalmol*. **21,** 486–491.

Carlson, N. B. and Kurtz, D. (2004) *Clinical Procedures for Ocular Examination*. Mc Graw Hill, New York.

Ciuffreda, K. J. (1992) Components of clinical near vergence testing. *J. Behav. Optom*. **3,** 3–13.

Ciuffreda, M. A., Ciuffreda, K. J. and Wang, B. (2006) Repeatability and variability of near vergence ranges. *J. Behav. Optom*. **17,** 39–46.

D'Augostino, R. B. and Stevens, M. A. (1986). *Tests for the Normal Distribution*. Marcel Dekker, New York.

Daum, K. M. (1991) Vergence amplitude. In: *Clinical Procedures in Optometry* (eds J. B. Eskridge, J. F. Amos and J. D. Bartlett), Lippincott, Philadelphia, pp. 91–98.

Feldman, J. M., Cooper, J., Carniglia, P., Schiff, F. M. and Skeete, J. N. (1989) Comparison of fusional ranges measured by Risley prisms, vectograms and computer orthopter. *Optom. Vis. Sci*. **66,** 375–382.

Goss, D. A. (1995) Effect of test sequence on fusional vergence ranges. *N. Engl. J. Optom*. **47,** 39–42.

Morgan, M. (1944) Analysis of clinical data. *Am. J. Optom. Arch. Am. Acad. Optom.* **21,** 477–491.

Penisten, D. K., Hofstetter, H. W. and Goss, D. A. (2001) Reliability of rotary prism fusional vergence ranges. *Optometry* **72,** 117–122.

Rosenfield, M., Ciuffreda, K. J., Ong, E. and Super, S. (1995) Vergence adaptation and the order of clinical vergence range testing. *Optom. Vis. Sci.* **72,** 219–223.

Rouse, M. W., Borsting, E. and Deland, P. N. (2002) Reliability of binocular vision measurements used in the classification of convergence insufficiency. *Optom. Vis. Sci.* **79,** 254–264.

Saladin, J. J. (1998) Phorometry and stereopsis. In: *Clinical Refraction* (ed. W. J. Benjamin), Saunders Company, Philadelphia, pp. 724–773.

Scheiman, M. and Wick, B. (2002). *Clinical Management of Binocular Vision. Heterophoric Accommodative and Eye Movement Disorders*. Lippincott Williams & Wilkins, Philadelphia.

Scheiman, M., Herzberg, H. and Frantz, K. (1989) A normative study of step vergence in elementary schoolchildren. *J. Am. Optom. Assoc.* **60,** 276–280.

Scheiman, M., Mitchell, G. L., Cotter, S., Cooper, J., Kulp, M., Rouse, M., Borsting, E., London, R. and Wensveen, J. (2005) A randomized clinical trial of treatment for convergence insufficiency in children. *Arch. Ophthalmol.* **123,** 14–24.

Schor, C. M. and Ciuffreda, K. J. (1983). *Vergence Eye Movements. Basic and Clinical Aspects*. Butterworth, Boston.

Sheedy, J. E. and Saladin, J. J. (1977) Phoria, vergence and fixation disparity in oculomotor problems. *Am. J. Optom. Physiol. Opt.* **54,** 474–478.

Sheedy, J. E. and Saladin, J. J. (1983) Validity of diagnostic criteria and case analysis in binocular vision disorders. In: *Vergence Movements: Basic and Clinical Aspects* (eds C. M. Schor and K. J. Ciuffreda), Butterworth, Boston, pp. 517–540.

Tannen, B. and Ciuffreda, K. J. (1995). *Eye Movement Basics for the Clinician*. Mosby Year-Book, St Louis.

Wesson, M. D. (1982) Normalization of prims bar vergences. *Am. J. Optom. Physiol. Opt.* **59,** 628–633.

Zadnik, K., Mutti, D. O. and Adams, A. J. (1992) The repeatability of measurement of the ocular components. *Invest. Ophthal. Vis. Sci.* **33,** 2325–2333.

Author Query Form

Journal: OPO

Article: 583

Dear Author,
During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

Query reference	Query	Remarks
1	**Production Editor: Please supply received, revised and accepted dates.**	
2	**Au: Please check and approve the changes made to the sentence:** 'Of the initial ... were conducted.	
3	**Au: Please define: TNO if applicable.**	
4	**Au: Please check this website address and confirm that it is correct.**	
5	**Au: Please check the volume number in reference Bland and Altman (1986).**	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	Ⓥ
Insert in text the matter indicated in the margin	⋏	New matter followed by ⋏ or ⋏ⓧ2
Delete	/ through single character, rule or underline or ⊢⊣ through all characters to be deleted	ꝺ or ꝺⓧ2
Substitute character or substitute part of one or more word(s)	/ through letter or ⊢⊣ through characters	new character / or new characters /
Change to italics	— under matter to be changed	
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	= under matter to be changed	=
Change to bold type	~ under matter to be changed	~
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	/ through character or ⋏ where required	under character e.g.
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	⊢-⊣
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	/ through character or ⋏ where required	Y
Reduce space between characters or words	| between characters or words affected	T