AOSIS

# Comparative analysis of Modified Thorington to the prism cover, von Graefe and Maddox rod tests

CrossMark

**Authors:**
Thokozile I. Metsing[1]
Solani D. Mathebula[2]

**Affiliations:**
[1]Department of Optometry, Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa

[2]Department of Optometry, Faculty of Health Sciences, University of Limpopo, Polokwane, South Africa

**Corresponding author:**
Thokozile Metsing, ingridm@uj.ac.za

**Background:** Heterophoria measurement is important in the evaluation of binocular vision. There are many different methods of heterophoria measurement. There are various methods of heterophoria measurements, and studies have compared different methods, some reporting little difference between the methods while others reported that measures can be significantly different.

**Aim:** The purpose of this study was to compare heterophoria measurements obtained using four clinical methods, that is, the modified Thorington test (MTT) on the Bernell Muscle Imbalance Measure (MIM) with those obtained from the prism cover test (PCT), Von Graefe test (VGT) and Maddox rod test (MRT), and thereby to establish their interchangeability.

**Setting:** The study was performed at an optometric clinic within a South African university.

**Methods:** Dissociated near horizontal heterophoria was measured on a sample of 30 optometry students (ages 19–25 years) using the PCT, VGT, MRT and MTT. Three measurements for each of the four tests per participant were obtained, and thereafter the means were applied in analysis of the results. Agreement between paired tests was assessed using Bland–Altman plots.

**Results:** The means for near or proximal heterophoria obtained using the four tests were $-2.63 \pm 1.9^\Delta$, $-2.90 \pm 1.7^\Delta$, $-3.37 \pm 2.0^\Delta$ and $-2.47 \pm 1.9^\Delta$, with a strong correlation for each pair ($p = 0.00$) (minus signs indicate exophoria [XOP]). Among the different combinations of tests, the MTT and PCT showed the best agreement, although the reliability was good for all four tests.

**Conclusion:** The measurement of near heterophoria can best be quantified either using the MTT or PCT because they showed the greatest level of agreement.

**Keywords:** Bernell muscle imbalance card; cover test; dissociated phoria; heterophoria; Maddox rod; modified Thorington test; Von Graefe test.

## Introduction

To see objects both focused and fused, a normal functioning binocular vision system requires a balanced interaction between sensory and motor components.[1,2,3,4] While the sensory component or fusion unifies the perception of images of the two eyes, motor fusion guarantees proper alignment of the eyes in such a manner that the sensory component can be maintained. If an anomaly is present in either of the two components, the functioning of the other can be significantly affected, resulting in ocular accommodative and nonstrabismic binocular vision (vergence) disorders.[3] During a clinical evaluation of the visual system, it is vital to evaluate and diagnose accommodative and nonstrabismic binocular vision disorders. One of the steps in the evaluation and management of binocular vision disorders is the measurement of the patient's heterophoria, usually at far (6.0 m) and near (0.4 m) working distances.

Heterophoria (or phoria) is a relative deviation of the visual axes that may appear in most individuals if one eye is artificially excluded from participating in vision,[5,6,7] therefore resulting in the suspension of the sensory and motor components of binocular vision. Heterophoria may be defined as a misalignment in the horizontal, vertical or cyclo-directions that is corrected by fusional reserves or disparity vergence.[8] Heterophoria is compensated for by fusional vergences through the mechanisms that involve both sensory and motor fusion. Esophoria (SOP) is present when the visual axes of each eye cross at a point in front of the object of regard or target, while exophoria (XOP) occurs when the visual axes intersect beyond the object of regard when all

stimuli to fusion is interrupted or absent.[6] Orthophoria occurs when the visual axes meet at the object of regard.[7] The magnitude of heterophoria is expressed in prism dioptres ([Δ] [prism dioptre]). Patients with decompensated heterophoria may experience symptoms such as headache, eye strain, blurred vision and diplopia that may affect visual efficiency, which is important for near vision activities such as reading and working on digital devices.[9]

There are several procedures available to clinicians for the subjective (or objective) measurement of heterophoria. These include the prism cover test (PCT), Von Graefe test (VGT), Maddox rod test (MRT) and modified Thorington test (MTT). Although these tests have some common features and are all used in clinical practice, they may differ in their measurements.[5,6,7,8,9,10,11,12,13,14,15,16,17,18,19] Factors that may affect heterophoria measurement include the technique used for dissociation, ability to control accommodation adequately, the length of time that fusion is suspended, the method by which heterophoria is quantified and the level of proximal convergence induced.[2] Each test presents different constraints that must be considered when used to assess heterophoria. However, several authors have found that the alternating cover test with prism neutralisation provides excellent repeatability, both within and between examiners.[19,20] Anderson et al.[20] also found that the cover test is a reliable measure of eye deviation, even when examiners are inexperienced.

The cover test is one of the most common methods to measure heterophoria in clinical practice. As it is an objective method, its results do not rely as heavily on the responses of the patient but rather on the ability and skills of the examiner to detect eye movements.[20,21,22,23,24] Some authors[20,21,22,23,24] have shown that the minimum amount of detectable eye movements under ideal condition using the PCT is approximately 2[Δ]. However, some authors have found no clinically relevant difference in the means between experienced and novice examiners.[23,24] Another possible limitation for the PCT is the use of different criteria for the neutralisation point. This is the source of interexaminer variability. The endpoint for the absence of eye movement that should be recorded for the test remains unclear. One possible endpoint is the first amount of prism with which no movement is seen (first neutral), and the other is when the prism causes an opposite movement of the eyes (reversal point), and the time of occlusion has a direct influence on the measured heterophoria.[6,20]

The VGT is very popular and is probably the usual or most common method for measuring heterophoria, where dissociation is achieved with the use of horizontal and vertical prisms.[14,25] The VGT is a subjective test that largely depends on the patient's responses. Several authors compared the repeatability of VGT with other methods and concluded that the method is less repeatable compared with PCT or MTT.[10,11,12,13,14,15,16,17,18] It was shown that the PCT yields lower phoria values than the VGT, while others reported that the VGT yields higher esophoric values than the PCT.[26,27,28] Other authors reported that the differences between the PCT and the VGT increases as mean horizontal heterophoria increases, for both distance and near vision in nonpresbyopic subjects.[28] Canto-Cerdan et al.[16] showed in presbyopic subjects that the PCT and VGT have a high level of agreement for both distance and near vision. However, the level of agreement is very low in nonpresbyopic subjects. Although there is controversy about the most repeatable test, it is agreed that the PCT and MTT offer better repeatability compared with the VGT.[12,21]

Several studies[12,15,16] have examined dissociated heterophoria tests but varied in the clinical techniques employed and in methods for statistical analysis. Heterophoria measurements has been a subject for many investigations, but uncertainty still exists about the most repeatable test. Several studies concluded that different methods for measuring heterophoria are not interchangeable because of low levels of agreement.[12,15,16] However, other studies found the MTT to be more repeatable compared with the VGT and MRT. Hence, this study was designed to compare results obtained using the MTT with those obtained using the PCT, VGT and MRT in the assessment of near horizontal heterophoria. In this study, the reliability and agreement as to whether this agreement fell within the prior limit of 2[Δ], which is generally regarded as the minimum detectable eye movement on the cover test, were also examined.

## Methods

Thirty ($n = 30$) nonpresbyopic university students (15 female and 15 male) participated in the study. The mean age and standard deviation were 22.9 ± 2.6 years, and ages ranged from 19 to 25 years. The inclusion criteria consisted of participants whose visual acuity was 6/6 (20/20) or better in each eye at 6.0 m and 0.4 m (40 cm), with no ocular motility disorders, strabismus, amblyopia, nystagmus or history of eye injury, eye diseases, visual therapy or refractive surgery. Participants with any history of medication for systemic or ocular diseases were excluded from the study. The screening tests included the cover-uncover test (unilateral cover test) to rule out the existence of strabismus (heterotropia or simply tropia) at far and near. Visual acuities, including the refractive status, were determined by the researchers. The visual acuity was measured using the Snellen chart at both distance and near. Autorefractor measurements were taken to screen for refractive errors, and participants included in the study were those with the spherical equivalence refractive error (SERE) of ± 0.50 dioptres (D).

Heterophorias were measured using four different methods, namely the PCT, VGT, MRT and MTT. Each of the four methods were performed by a different examiner with experience in heterophoria measurement procedures, but none of them were aware of the results obtained from the other examiners. This prevented extraneous influences on the various examiners, and the results for one examiner were not influenced by those of the others. Three

measurements for all heterophoria tests per participant were performed and the mean recorded for subsequent analysis.

## Prism cover test

The PCT was performed to evaluate the heterophoria with the eyes alternately occluded and eye movements were observed. Each participant was asked to hold a visual acuity card at 40 cm and fixate at a 6/9 (20/30) visual acuity line. A minimum occlusion time of 5 s per eye was used to minimise the effect of vergence adaptation. A prism bar held at 1 cm from the spectacle plane with powers of $1^\Delta$, $2^\Delta$, $4^\Delta$ to $20^\Delta$ in $2^\Delta$ steps and powers of $25^\Delta – 45^\Delta$ in $5^\Delta$ steps was used to neutralise the eye movements. The amount of prism power that neutralised the eye movements was regarded as the measurement for horizontal heterophoria.

## Von Graefe test

The phoropter Risley rotary prisms were used for measuring heterophoria using the Von Graefe test. A dissociating prism of $6^\Delta$ base-up was placed in front of the left and a measuring prism of $12^\Delta$ base-in before the right eye. Participants were instructed to fixate at the 6/6 (20/20) line of the near visual acuity card. Participants were instructed to fixate at the lower target (the nonmoving target) and keep the letters clear at all the times. The magnitude of the $12^\Delta$ base-in was altered in one-prism dioptre steps until the participant reported that the two images were vertically aligned. The amount of the horizontal prism that brought the diplopic images into vertical alignment was recorded as the horizontal heterophoria. Three measurements per participant were performed and the means were recorded.

## Modified Thorington test

The MTT is a mainly subjective test to measure heterophoria. It uses the Bernell Muscle Imbalance Measure (MIM) card (Bernell Corp., Indiana) with a column of numbers separated by one prism dioptre held at 40 cm. A penlight torch was shown to the participant through a hole in the centre of the MIM card while they held a Maddox rod lens horizontally oriented in front of the right eye. Participants were asked to report the number through which the red line (created by the Maddox rod) passed on the MIM card. The MIM card used had a measurement range from $28^\Delta$ SOP to $28^\Delta$ XOP, with a resolution of one prism dioptre. Each participant was asked to report the number on the card through which the red vertical line appeared to be on the lateral scale. Numbers to the right of the light indicated the presence of an SOP and those to the left indicated an XOP.

## Maddox rod test

Heterophoria was measured at 40 cm with a red Maddox rod lens in front of the right eye and a measuring prism, which was a prism bar from $1^\Delta$ to $45^\Delta$ prism in front of the left eye. A spot of light was held at 40 cm. This caused the participant to see a vertically oriented red line and spot of light. The examiner then requested the participant to report when the line and spot of light were superimposed. The measuring prism was altered in one prism dioptre steps until superimposition occurred. The measurements were performed three times, and the mean of the measurements obtained was recorded and used for the analysis of the findings. If superimposition of the white spot of light and the red streak of light was reported from the onset of the test, then the findings were recorded as orthophoric.

## Statistical analysis

Statistical analyses were performed using the Statistical Package for Social Sciences (SPSS, Inc., Chicago, IL, United States). The Shapiro–Wilk test was used to check if each heterophoria measurements were normally distributed. Tests were considered statistically significant if $p < 0.05$. For this study, XOP was represented with a negative sign and SOP with a positive sign. Box plots (see Figure 1) were used to compare results across the four methods.
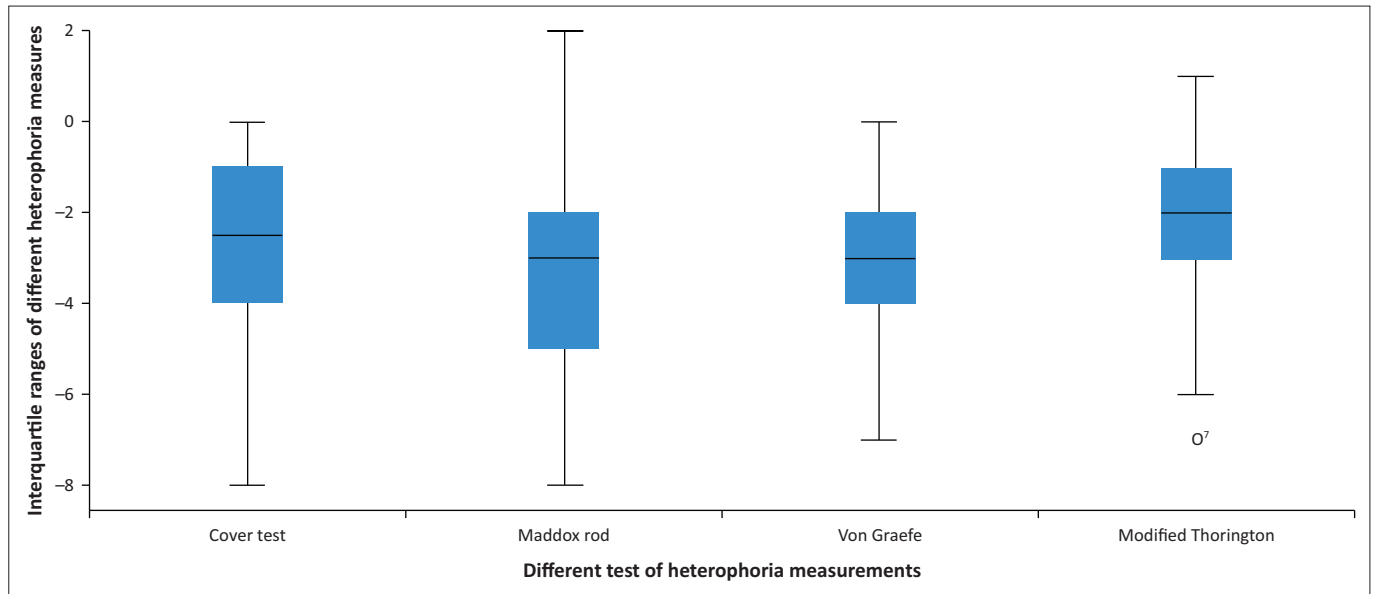
Bland–Altman analysis was performed to determine the level of agreement.[29] The 95% limits of agreements (LoA) were presented by the upper and lower lines (see Figure 2), which are equal to the mean difference (*bias*) ± 1.96 standard deviations. The middle line represented the mean difference. Based on the minimum detectable eye movements, the maximum acceptable 95% limits were defined here as ± $2^\Delta$. The $2^\Delta$ was based on the minimum detectable eye movements on cover test. Intraclass correlation coefficients (ICCs) were determined for paired methods to assess the reliability using a two-way mixed absolute agreement model.[30] The magnitude of the ICC was interpreted according to the levels of reliability: less than 0.50 was regarded as poor reliability (agreement), moderate reliability was 0.50–0.75 and good reliability was equal to or more than 0.75.
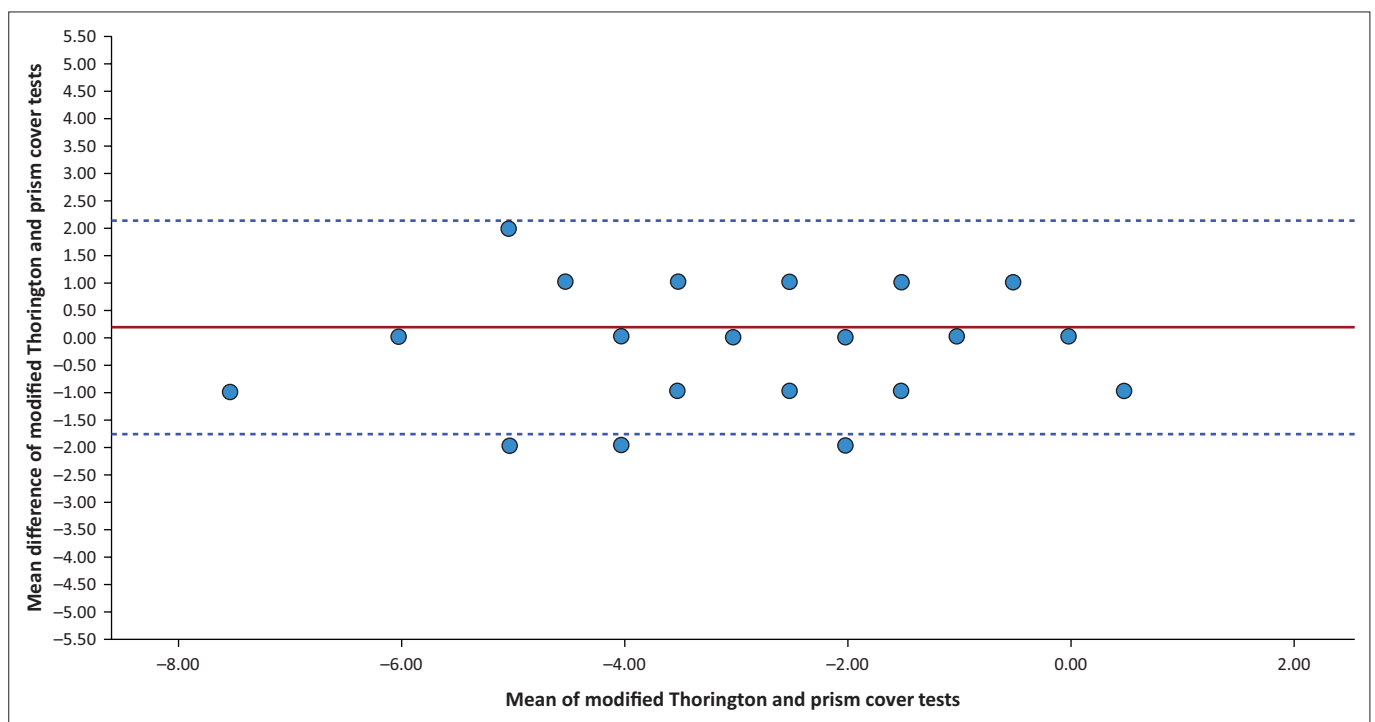
## Ethical considerations

Ethical clearance to conduct this study was obtained from the University of Johannesburg Health Science Research Ethics Committee (ref. no. 01-24-2019). The study adhered to the Declaration of Helsinki. All participants were informed about the nature of the study before providing informed consent.

# Results

Table 1 shows the tests of normality. In SPSS two statistical tests of normality, namely the Kolmogorov–Smirnov (K–S) test and the Shapiro–Wilk test are available. Although provided here, the output from the K–S test was not used because of its low power. The Shapiro–Wilk test is the best choice for testing the normal distribution for a sample size

**FIGURE 1:** Boxplots showing the interquartile ranges for four near heterophoria methods. Positive and negative values represent esophoria and exophoria, respectively. Measurements were expressed in prism dioptres ($^\Delta$).



**FIGURE 2:** Bland–Altman plot showing the agreement between modified Thorington test and prism cover test. The central red line indicates the mean difference (= 0.17$^\Delta$ in Table 4) between the measurements. Observing the graph, many points are located relatively close to the red solid line that represents the mean difference. The two outer lines indicate the upper and lower limits of the 95% limits of agreement (LoA) interval. The smaller this region and the closer to zero, the greater is the level of agreement between the two samples concerned. The units for the horizontal and vertical axes are prism dioptres.

less than 50 participants.[31] As all variables studied presented normal distributions ($p > 0.05$, as per Shapiro–Wilk test), parametric tests for statistical analyses determined any statistically significant differences between methods.

Table 2 shows the descriptive statistics for the 30 participants and allows comparison of the samples for PCT, VGT, MRT and the MTT. The mean value with the MRT was $-3.37 \pm 2^\Delta$, and if one considers magnitudes only, this is slightly greater (more exophoric) as compared with

those for MTT, PCT and MRT with means of $-2.47 \pm 1.9^\Delta$, $-2.63 \pm 1.9^\Delta$ and $-2.90 \pm 1.7^\Delta$, respectively. There was, however, no statistically significant difference between these mean values, $p > 0.05$.

Figure 1 shows four boxplots, each of which has the median as a bold horizontal line inside each box and the IQR (between 25% and 75% percentiles) as the length of each box. The whiskers are the lines extending from the top and bottom of each box, representing the minimum and

maximum values shown in Table 2. The whiskers are within 1.5 times the IQR from either end of each box concerned. Normally distributed data have a median line at approximately the centre of the box with symmetric whiskers. The medians were $-2.5^\Delta$ (PCT), $-3^\Delta$ (MRT), $-3^\Delta$ (VGT) and $-2^\Delta$ (MTT), respectively. Measurements greater than 1.5 times the IQR are considered outliers. In Figure 1, the measurement for participant number 7 with MTT could be considered as an outlier.

**TABLE 1:** Tests for data normality for four tests for near horizontal heterophoria.

| Heterophoria tests | Kolmogorov–Smirnov | | Shapiro–Wilk | |
| --- | --- | --- | --- | --- |
| | Statistic | p | Statistic | p |
| Prism cover test (PCT) | 0.158 | 0.053 | 0.93 | 0.149 |
| Von Graefe test (VGT) | 0.172 | 0.023 | 0.95 | 0.192 |
| Maddox rod test (MRT) | 0.177 | 0.018 | 0.94 | 0.105 |
| Modified Thorington test (MTT) | 0.184 | 0.011 | 0.92 | 0.133 |

Note: A value for $p > 0.05$ indicates univariate data normality, and this was true for all tests with the Shapiro–Wilks but not with the Kolmogorov–Smirnov tests.

**TABLE 2:** Descriptive statistics for the four near heterophoria methods.

| Statistics | Modified Thorington test (MTT) | Prism cover test (PCT) | Von Graefe test (VGT) | Maddox rod test (MRT) |
| --- | --- | --- | --- | --- |
| Mean ± s.d. | $-2.47 \pm 1.9$ | $-2.63 \pm 1.9$ | $-2.90 \pm 1.7$ | $-3.37 \pm 2.0$ |
| Median | $-2.00$ | $-2.50$ | $-3.00$ | $-3.00$ |
| Minimum | $-7$ | $-8$ | $-7$ | $-8$ |
| Maximum | 1 | 0 | 0 | 2 |
| Skewness | $-0.7$ | $-0.8$ | $-0.6$ | 0.2 |
| Kurtosis | 0.2 | 0.8 | 0.1 | 1.2 |
| IQR | | | | |
| 25% | $-3.25$ | $-4.50$ | $-4.00$ | $-5.00$ |
| 50% | $-2.00$ | $-2.50$ | $-3.00$ | $-3.00$ |
| 75% | $-1.00$ | $-1.00$ | $-2.00$ | $-2.00$ |

Note: Units are in prism dioptres ($^\Delta$). Positive values represent esophoria and negative exophoria. For univariate normality, skewness and kurtosis should be zero.

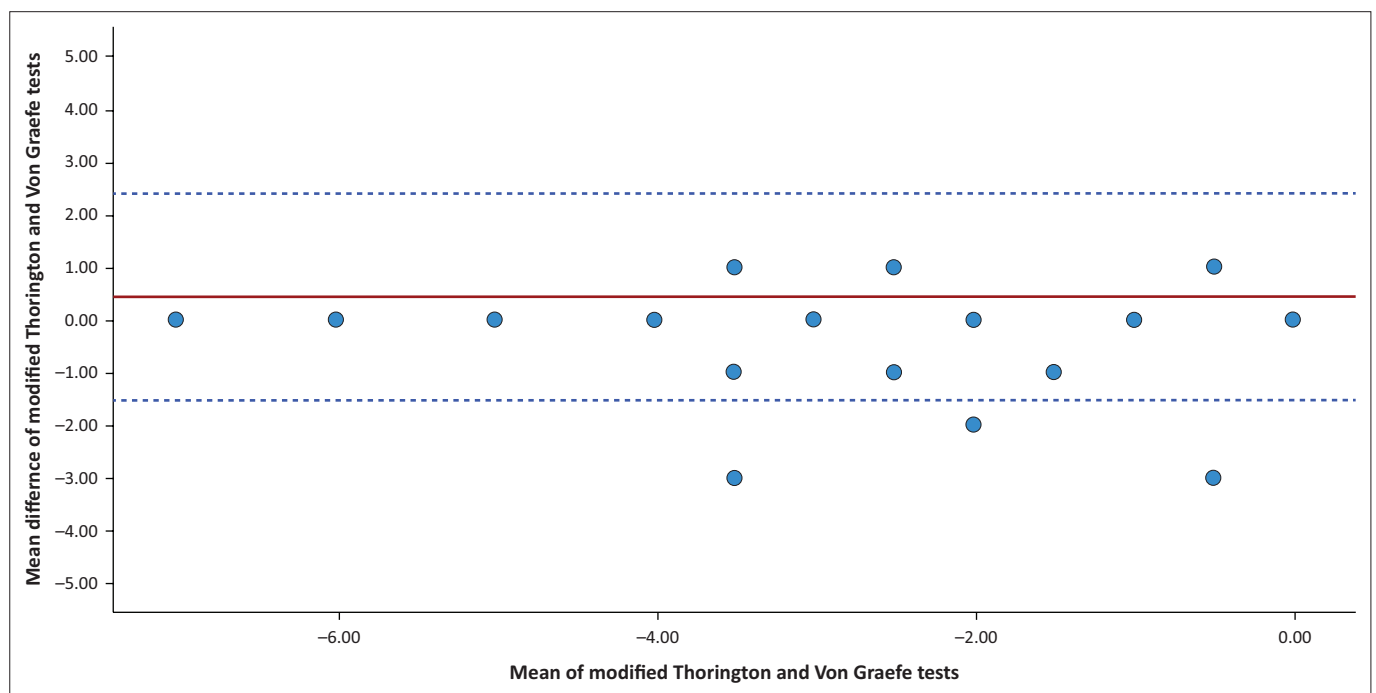s.d., standard deviation; IQR, interquartile range.

A correlation analysis (Table 3) was performed and correlations between means of paired methods ranged from 0.54 (moderate) to 0.86 (strong).

Some of the Bland–Altman analysis is indicated in Table 4 and Figure 2, Figure 3 and Figure 4. The mean differences (*biases*) between paired methods are indicated in Table 4. The correlations between difference for MTT and PCT, VGT and PCT and between VGT and MRT were not statistically significant ($p > 0.05$) at 95% confidence level. However, those obtained between MTT and VGT, MTT and MRT and between PCT and MRT were statistically significant, $p < 0.05$. Mean differences below $2^\Delta$ are too small to be considered clinically significant despite being statistically significant, with the $p$-value $< 0.05$ at 95% confidence level. Bland–Altman plots for the different tests were obtained, although only three (Figure 2, Figure 3 and Figure 4) are provided here. As the scales for the horizontal and vertical axes are the same, agreement between the different methods could be easily compared. The Bland–Altman analysis showed that the mean difference was small ($< 2^\Delta$) for all pairs of methods used for heterophoria measurement. However, all LoA fell outside the predefined criterion of $< 2^\Delta$. The narrower (see widths in

**TABLE 3:** Correlation analysis for four near horizontal heterophoria methods.

| Paired methods | Correlation (r) | p |
| --- | --- | --- |
| MTT versus PCT | 0.86 | 0.00 |
| MTT versus VGT | 0.85 | 0.00 |
| MTT versus MRT | 0.77 | 0.00 |
| PCT versus VGT | 0.64 | 0.00 |
| PCT versus MRT | 0.74 | 0.00 |
| VGT versus MRT | 0.54 | 0.00 |

MTT, modified Thorington test; PCT, prism cover test; VGT, Von Graefe test; MRT, Maddox rod test.



**FIGURE 3:** Bland–Altman plot of means versus difference of modified Thorington test and Von Graefe test measurements. The solid horizontal red line indicates the mean difference (= $0.43^\Delta$) and the limits of agreement are represented by the upper and lower dotted lines.

Table 4) the 95% LoA, the better the agreement. The best agreement was between MTT and PCT, as indicated by a small mean difference of only 0.17$^\Delta$, and the width of the LoA at 3.92$^\Delta$, although the PCT tended to give less esophoric measurements. The pair that showed the worst agreement with the MTT was MRT (0.90 ± 1.3$^\Delta$). The mean difference was, however, still very small (< 1$^\Delta$), but the width of the LoA was larger at 5$^\Delta$, so essentially the two methods could still be used interchangeably in clinical practice, although there would be some individuals for whom the two methods might not be in very close agreement.

Table 5 shows the ICC produced for the different comparisons. Intraclass correlation coefficients correlate the size of the measurement error to the variability in true values between participants and take on values between 0 (representing unreliable) and 1 (indicating perfect reliability). If the reliability is high and close to one, measurement errors are small in comparison to the variability between participants and differences between measurements of two participants; therefore, errors occur because of the difference in their true values rather than the measurement error.[30] Firstly, the relationship between all four tests was tested, and secondly the MTT was compared with the other tests. All ICC indicated good consistency or agreement, and ICC (= 0.9) were statistically significant, $p < 0.009$ (see Table 5).
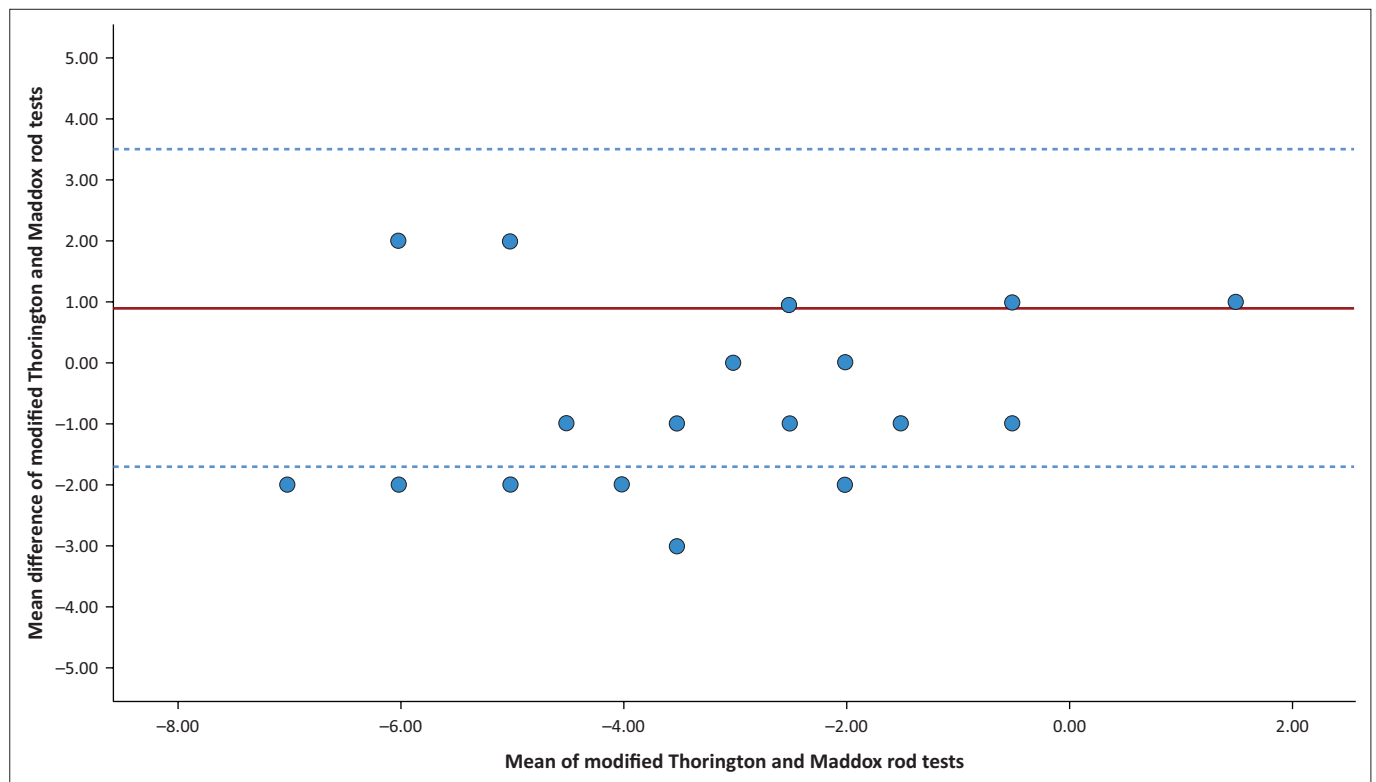
## Discussion

The purpose of this study was to compare near dissociated horizontal heterophorias obtained using the MTT with the PCT, VGT and MRT. The MTT was investigated against the other three commonly used methods because previously, MTT was found to have the highest interexaminer

**TABLE 4:** Mean differences of heterophoria measures in prism dioptres with 95% reproducibility limits (± 1.96√2SD ≈ ± 2.77SD) and 95% limits of agreement.

| Comparisons | Mean ± s.d. | 95% reproducibility limits | 95% confidence intervals for LoA | | | p |
|---|---|---|---|---|---|---|
| | | | Lower limit | Upper limit | Width | |
| MTT-PCT | 0.17 ± 1.0 | ± 2.77 | −2.13 | 1.79 | 3.92 | 0.4 |
| MTT-VGT | 0.43 ± 1.0 | ± 2.77 | −2.39 | 1.53 | 3.92 | **0.0** |
| MTT-MRT | 0.90 ± 1.3 | ± 3.60 | −3.40 | 1.60 | 5.00 | **0.0** |
| VGT-PCT | −0.27 ± 1.6 | ± 4.43 | −3.40 | 2.83 | 6.23 | 0.4 |
| VGT-MRT | 0.47 ± 1.8 | ± 4.99 | −3.03 | 4.00 | 7.03 | 0.2 |
| PCT-MRT | −0.73 ± 1.4 | ± 3.88 | −3.40 | 2.00 | 5.40 | **0.0** |

Note: Positive and negative values represent esophoria and exophoria, respectively. The positive sign indicates that the mean for the first test is more esophoric (or less exophoric) than the mean for the second test in any comparison, while the negative sign shows the opposite. For example, on average MTT gave slightly more esophoric results than for PCT (see first comparison). In the last column, significant results are indicated with bold text.
MTT, modified Thorington test; PCT, prism cover test; VGT, Von Graefe test; MRT, Maddox rod test.

**TABLE 5:** Inter-rater correlation coefficients for modified Thorington test, prism cover test, Von Graefe test and Maddox rod test.

| Comparisons | ICC | 95% confidence intervals for ICC | | F test with true value | | | |
|---|---|---|---|---|---|---|---|
| | | Lower bounds | Upper bounds | Values | df1 | df2 | p |
| All four tests | 0.9 | 0.85 | 0.96 | 11.83 | 29 | 87 | 0.00 |
| MTT and PCT | 0.9 | 0.84 | 0.96 | 12.95 | 29 | 29 | 0.00 |
| MTT and VGT | 0.9 | 0.82 | 0.96 | 11.68 | 29 | 29 | 0.00 |
| MTT and MRT | 0.87 (≈ 0.9) | 0.72 | 0.93 | 7.62 | 29 | 29 | 0.00 |

MTT, modified Thorington test; PCT, prism cover test; VGT, Von Graefe test; MRT, Maddox rod test.



**FIGURE 4:** Bland–Altman plot showing the agreement between modified Thorington test and Maddox rod test. Although most points are within the limits of agreement, there are several points outside the 95% limits.

repeatability.[12,32] The VGT and MRT procedures are commonly used methods for subjective dissociated heterophoria testing.[14,25] In this study, one examiner was used for each test, and each examiner was masked to the other test results to avoid any possible single examiner bias across methods.

The mean values or magnitudes at near 0.4 m obtained for MTT, PCT, VGT and MRT were 2.47 ± 1.9$^\Delta$ XOP, 2.63 ± 1.9$^\Delta$ XOP, 2.90 ± 1.7$^\Delta$ XOP and 3.37 ± 2.0$^\Delta$ XOP, respectively, so approximately 2.5$^\Delta$ – 3.4$^\Delta$ across methods. For near vision, other authors established means between 3$^\Delta$ and 4$^\Delta$ of XOP, using the VGT in nonpresbyopic subjects.[12,15,16,18] In this study, the means obtained using the MTT, PCT and VGT are slightly less exophoric than those obtained using the MRT (see Table 2 and Figure 1). The findings of this study imply that, on average, the VGT and MRT are slightly more exophoric when compared with the MTT and PCT findings. These results are consistent with those from Rainey et al.,[12] who also reported more exophoric means when using the VGT. The results of this study are also in agreement with those of Calvin et al.,[27] who found that the PCT tended to yield less XOP than the VGT.

The mean ± s.d. near heterophoria determined using the MTT in this study was 2.47$^\Delta$ ± 1.9$^\Delta$ XOP and 2.63$^\Delta$ ± 1.9$^\Delta$ XOP for the PCT. Among the different mean differences, the MTT and the PCT showed the least variability, which indicates best agreement. The mean difference between MTT and PCT (0.17$^\Delta$ ± 1.0$^\Delta$) was not statistically significant ($p$ = 0.4) at the 95% confidence level. Mean differences below 1$^\Delta$ are so small that they may be considered clinically insignificant, although sometimes being statistically significant. The mean differences provided by the different techniques differed by less than 0.5$^\Delta$ with the exceptions of MTT versus MRT and PCT versus MRT. Table 3 shows greater variation in heterophoria measures for the MTT and MRT, although probably clinically insignificant. Both methods are based on the same manner of dissociation that creates rivalry between the two eyes, where the right eye sees a vertical red streak or line and the left eye views a spot of light. The main reason for the differences in mean heterophoria measurement could be the extent of dissociation provided by the techniques. As participants wore their distance refractive compensation, the role played by ocular accommodation could probably be regarded as minimal.

Rainey et al.[12] and Wong et al.[32] reported their mean near MTT findings to be 1$^\Delta$ – 2$^\Delta$ less exophoric than for their VGT finding. The present study's results showed the mean MTT to be 0.4$^\Delta$ less exophoric as compared with the VGT mean. The mean differences (0.43$^\Delta$ ± 1$^\Delta$) between MTT and VGT were statistically significant ($p$ = 0.00) at the 95% level confidence level. Similarly, Maples et al.[10] reported a significant difference between the MTT and VGT ($p$ < 0.01). This study's results for MTT were dissimilar to Maples et al.[10] Their study reported that the mean near heterophoria with MTT was 2.09$^\Delta$ XOP, while the mean heterophoria at near with VGT was 6.33$^\Delta$ XOP, which was dissimilar to the results of the this study (see

Table 2). Studies by Andrew et al.[33] and Hyun-Jin et al.[34] asserted that the differences in heterophoria measurements were not significant, but the measurements using a phoropter showed a greater coefficient of variation. This was also reported by Goss et al.,[11] in that VGT showed more variability on near horizontal heterophoria measurement. The variation among different authors may be because of the possible biases in the sample, examiners and methods, and the instrument manipulation. Similarly, Faccin and Maffioletti[8] reported that the VGT shows more XOP and larger LoA at near than the MTT.

Bland–Altman plots were used to describe the agreement between MTT and the PCT, VGT and MRT using the LoA at the 95% level (mean difference ± 1.96 [s.d.]). The mean differences and the 95% LoA for the different heterophoria methods are shown in Table 3. The absence of significant differences in mean differences between MTT and PCT, VGT and PCT and VGT and MRT imply that there were no real differences in mean differences (*biases*) among the four heterophoria methods. As the mean differences are small (less than 1$^\Delta$), for clinical purposes the four techniques can be used interchangeably. Atuanya et al.[35] compared the MTT and VGT on near lateral heterophoria among 100 emmetropes. The near heterophoria measurement of VGT and MTT did not differ statistically, although VGT showed a broader range and more exophoric measurements than the MTT. Yu and Ha[36] investigated the difference in the value of horizontal heterophoria measured in 72 college students using the VGT and MRT. There were 21 orthophoric, 36 exophoric and 15 esophoric participants. The authors did not find definite differences for horizontal heterophoria by any test method. The results of this study for limits of reproducibility (see Table 4 where the greatest limit of reproducibility was ± 4.99$^\Delta$) were, however, slightly dissimilar to previous studies where Antona et al.[37] found ± 6.54$^\Delta$, Cebrian et al.[28] found ± 2.95$^\Delta$ and Canto-Cerdan et al.[16] found ± 6.74$^\Delta$.

Figures 2–4 illustrate the extent of agreement between the different heterophoria methods. The narrower the agreement intervals, the better the agreement. As the same scales were used for the horizontal and vertical axes, agreement between the tests can be easily compared. The best agreement was observed for MTT and PCT, as indicated by a small mean difference (0.17$^\Delta$) and small limit of reproducibility (± 2.77$^\Delta$). The worst agreement with MTT was MRT (± 3.60$^\Delta$), and VGT similarly showed poorer agreement with PCT (± 4.43$^\Delta$) and MRT (CoA = ± 4.99$^\Delta$). The pair of PCT with MRT was ± 3.88$^\Delta$. Goss et al.[11] reported ± 3.50$^\Delta$ between MTT and VGT. The results of this study showed that the MTT tended to give more esophoric values than the VGT. A possible explanation for the more exophoric values obtained using the VGT could be that the VGT starts with 12$^\Delta$ base-in, and the participants could show vergence disparity response to try to reduce the distance between images. As for the MRT procedure, the 12$^\Delta$ base-in was placed before the left eye when the white spot and the vertical red line were not overlapping.

Table 4 shows ICC, the reliability parameter, which is the correlation between any two measurements made on the same participants. If the reliability is high, measurement errors are small compared with the actual differences between participants. Reliability takes values between zero and one, with a value of one corresponding to zero measurement error and a value of zero meaning that all the variability in measurements is because of measurement error. As reliability is a dimensionless quantity, it is difficult to interpret and decide which value constitutes high reliability, as the decision is made subjectively. However, in this study all the ICC values were close to 0.9, which suggests reliability was good.

This study had some limitations; for example, the sample size (30 participants only) was small. The examiners in this study were final-year optometry students, and the participants were also final-year optometry students who understood the instructions and procedures. However, this does not mean this study's findings are not relevant to near horizontal heterophorias in the general population. Intra- and interexaminer and even participant variability may be large in clinical settings, as maintaining the target clarity may be difficult to control in untrained observers. Only three measures were averaged per method per participant, and this could perhaps be improved in future studies. However, participant fatigue also needs to be considered if too many measurements per method are obtained.

## Conclusion

It is recommended that the subjective measurement of near horizontal heterophoria best be quantified either using the MTT or PCT. Modified Thorington test and PCT have a high level of agreement, and both can be used interchangeably. The MTT is quick and simple to perform and easy for participants to follow instructions. The findings of this study show that the VGT and MRT are slightly less interchangeable with MTT for near horizontal heterophoria measurement.

## Acknowledgements

### Competing interests

### Authors' contributions

T.I.M. was responsible for the conceptualisation of the study and methodology, including taking part in the analysis of the collected research data and the writing of the original draft. S.D.M. was responsible for the analysis of the research data, reviewing and editing of the draft.

### Funding information

### Data availability

Data supporting the findings of this study are available from the corresponding author, T.I.M.

### Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

## References

1. Rutstein R, Daum K. Anomalies of binocular vision: Diagnosis and management. 1st ed. St Louis, MO: Mosby; 1997.

2. Evans BJW. Pickwell's binocular vision anomalies. 5th ed. London: Butterworth-Heinemann; 2021.

3. Scheiman M, Wick B. Clinical management of binocular vision; heterophoria, accommodative, and eye movement disorders. 5th ed. Philadelphia, PA: Lippincott William & Wilkins; 2015.

4. Von Noordern GK, Campos EC. Binocular vision and ocular motility. 6th ed. St Louis, MO: Mosby; 2002.

5. Hofstetter HW, Griffin JR, Berman MS. Dictionary of visual science and related clinical term. 5th ed. Boston, MA: Buttersworth-Heinemann; 2000.

6. Dowley D. Heterophoria. Optom Vis Sci. 1990;67(6):456–460. https://doi.org/10.1097/00006324-199006000-00010

7. Walline JJ, Mutti DO, Zadnik K. Development of phoria in children. Optom Vis Sci. 1998;75(8):605–610. https://doi.org/10.1097/00006324-199808000-00026

8. Facchin A, Maffioletti S. Comparison, within-session repeatability and normative data of three phoria tests. J Optom. 2021;14(3):263–274. https://doi.org/10.1016/j.optom.2020.05.007

9. Christian LW, Nandakumar K, Hrynchak PH, Irving EL. Visual and binocular status in elementary school children with a reading problem. L Optom. 2018;11: 160–166. https://doi.org/10.1016/j.optom.2017.09.003

10. Maples WC, Savoy RS, Harville BJ. Comparison of distance and near heterophoria by two clinical methods. Optom Vis Dev. 2009;40(2):100–106.

11. Goss DA, Reynolds JL, Todd RE. Comparison of four dissociated phoria tests: Reliability, and correlation with symptom survey scale. J Behav Optom. 2010;21(4):99–104.

12. Rainey BB, Schroeder TL, Goss DA. Inter-examiner repeatability of heterophoria tests. Optom Vis Sci. 1998;75(10):719–726. https://doi.org/10.1097/00006324-199810000-00016

13. Schroeder TL, Rainey BB, Goss DA. Reliability of and comparison among methods of measuring dissociated phoria. Optom Vis Sci. 1996;73(6):389–397. https://doi.org/10.1097/00006324-199606000-00006

14. Cassilas EC, Rosenfield M. Comparison of subjective heterophoria testing with a phoropter and trial frame. Optom Vis Sci. 2006;83(4):237–241. https://doi.org/10.1097/01.opx.0000214316.50270.24

15. Sanker N, Prabhu A, Ray A. A comparison of near-dissociated heterophoria test in free space. Clin Exp Optom. 2012;95(11):638–642. https://doi.org/10.1111/j.1444-0938.2012.00785.x

16. Canto-Cerdan M, Cacho-Martinez P, Garcia-Munoz A. Measuring the heterophoria: Agreement between two methods in non-presbyopic and presbyopic patients. J Optom. 2018;11(3):153–159. https://doi.org/10.1016/j.optom.2017.10.002

17. Anstice NS, Davidson B, Field B. The repeatability and reproducibility of four techniques for measuring horizontal heterophoria: Implications for clinical practice. J Optom. 2020;14(3):275–281. https://doi.org/10.1016/j.optom.2020.05.005

18. Goss DA, Moyer BJ, Teske MC. A comparison of dissociated phoria test findings with Von Graefe phorometry and modified Thorington testing. J Behav Optom. 2008;19(6):145–149.

19. Jackson TW, Goss DA. Variation and correlation of standard clinical phoropter tests of phorias, vergence ranges, and relative accommodation in a sample of school-age children. J Am Optom Assoc. 1991;62(7):540–547.

20. Anderson HA, Manny RE, Cotter SA, Mitchell GL, Irani JA. Effect of examiner experience and technique on the cover test. Optom Vis Sci. 2010;87(3):168–175. https://doi.org/10.1097/OPX.0b013e3181d1d954

21. Johns HA, Manny RE, Fern K, Hu Y-S. The intraexaminer and interexaminer repeatability of the alternate cover test using different prism neutralisation endpoints. Optom Vis Sci. 2004;81(12):939–946.

22. Hrynchak PK, Herriot C, Irving EL. Comparison of alternate cover test reliability at near in non-strabismus between experienced and novice examiners. Ophthalmic Physiol Opt. 2010;30(3):304–309. https://doi.org/10.1111/j.1475-1313.2010.00723.x

23. Fogt N, Baughman BJ, Good G. The effect of experience on the detection of small eye movements. Optom Vis Sci. 2000;77(12):670–674. https://doi.org/10.1097/00006324-200012000-00014

24. Ludvigh E. Amount of eye movement objectively perceptible to the unaided eye. Am J Ophthalmol. 1949;32(5):649–650. https://doi.org/10.1016/0002-9394(49)91415-4

25. Tsotetsi AL, Mathebula SD. Comparison of phoropter and trial frame-based Von Graefe heterophoria measures in non-presbyopic participants. Afr Vision Eye Health. 2021;80(1):a645. https://doi.org/10.4102/aveh.v80i1.645

26. Sparks BI. Phoria variation secondary to cover test technique at near. Optometry. 2002;73(1):51–54.

27. Calvin H, Rupnow P, Grosvenor T. How good is the estimated cover test at predicting the Von Graefe phoria measurement? Optom Vis Sci. 1996;73(11):701–706. https://doi.org/10.1097/00006324-199611000-00005

28. Cebrian JL, Antona B, Barrio A. Repeatability of the modified Thorington card used to measure far heterophoria. Optom Vis Sci. 2014;91(7):786–792. https://doi.org/10.1097/OPX.0000000000000297

29. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;327(8476):307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

30. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(20):155–163. https://doi.org/10.1016/j.jcm.2016.02.012

31. Razali NM, Wah YB. Power comparisons of Shapiro-Wilk, Kolmogov-Smirnov, Lilliefors and the Anderson-Darling tests. J Stat Model Anal. 2011;2(1):21–33.

32. Wong EPF, Fricke TR, Dinardo C. Interexaminer repeatability of a new, modified Prentice card compared with established phoria tests. Optom Vis Sci. 2002;79(6):370–375. https://doi.org/10.1097/00006324-200206000-00010

33. Andrew KC, Abby L, Jessie C, Wong K. Comparison of near heterophoria tests under varying conditions on an adult sample. Ophthalmic Physiol Opt. 2014;25(2):162–167. https://doi.org/10.1111/j.1475-1313.2005.00270.x

34. Hyun-Jin O, Ha-Young D, Seung-Jin O. A study on the measurement and tendency of heterophoria using Von Graefe test and Maddox rod test. J Digital Convergence. 2012;10(11):485–491.

35. Atuanya GN, Uchendu VC, Musa MJ, Akpalaba RUE. Comparative study between modified Thorington test and Von Graefe phorometry on near lateral phoria assessment among emmetropes. Nigerian Res J Engineering Environment Sci. 2020;5(2):694–699.

36. Yu D, Ha E. Comparison of phoria test among prism settings of Von Graefe technique. J Korean Ophthalmic Opt Soc. 2015;20(2):211–218. https://doi.org/10.14479/jkoos.2015.20.2.211

37. Antona B, Gonzalez E, Barrio A. Strabometry precision: Intra-examiner repeatability and agreement in measuring the magnitude of the angle of latent binocular ocular deviations (heterophoria or latent strabismus). Binocul Vis Strabolog Q Simms Romano. 2011;26(2):91–104.