Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clinical Validation of Virtual Reality–Based Oculography for Quantitative Ocular Alignment Assessment in Strabismus

Donghun Lee

eydonghun@gmail.com

Daegu Catholic University

Kimia Rezaei

University of California, Irvine

Mia Jia Wang

University of California, Irvine

Donny W. Suh

University of California, Irvine

Research Article

Keywords: Strabismus, Virtual reality, Oculography, Ocular alignment, Eye Tracking

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10922714/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

To evaluate the clinical validity of a virtual reality (VR)-based eye-tracking system for quantitative ocular alignment assessment by comparing device-derived measurements with conventional clinical measurements in strabismus.

Methods

We retrospectively reviewed medical records of consecutive patients undergoing strabismus evaluation. We assessed ocular alignment using conventional prism cover testing and a VR-based oculography system (neos™, machineMD, Bern, Switzerland). We compared total deviation measured by alternate prism cover testing (APCT) and manifest deviation measured by simultaneous prism cover testing (SPCT) with device-derived measurements. We assessed associations using Spearman correlation and agreement using Bland–Altman analysis. Subgroup analyses were performed by horizontal deviation magnitude (<15 vs. ≥15 prism diopters [PD]).

Results

Forty-six patients (mean age, 37.61 ± 20.41 years; 25 [54.3%] female) were included. Device-derived measurements correlated significantly with conventional measurements for horizontal total deviation (Spearman ρ = 0.780, $P < 0.001$), vertical total deviation (ρ = 0.728, $P < 0.001$), horizontal tropia (ρ = 0.783, $P < 0.001$), and vertical tropia (ρ = 0.424, $P = 0.003$). Bland–Altman analysis showed closer agreement for horizontal than vertical measurements. Measurement differences did not differ significantly between the <15 PD and ≥15 PD groups, although greater variability was observed in the ≥15 PD group.

Conclusions

VR-based oculography demonstrated significant associations with conventional measurements, with better agreement for horizontal than vertical deviations. Greater variability with larger horizontal deviations suggests cautious interpretation in large-angle strabismus.

Key Message

What is known?

Conventional prism cover testing remains the clinical standard for quantitative assessment of ocular alignment, while evidence supporting virtual reality–based oculography in patients with strabismus remains limited.

What does this study add?

Virtual reality–based oculography demonstrated significant correlations with conventional ocular alignment measurements, with stronger agreement for horizontal than vertical deviations. Individual-level measurement variability was greater in patients with horizontal deviations ≥15 prism diopters, supporting cautious interpretation in large-angle strabismus.

Introduction

Strabismus was defined as constant or intermittent heterotropia of any dimension at distance or near target fixation with refractive error correction.[1,2] Strabismus can be both a cause and a consequence of impaired binocularity, including reduced binocular vision, stereopsis, and visual development.[3] In particular, strabismic amblyopia has been reported to exhibit more significant peripheral binocular imbalance than other amblyopia subtypes.[4] In addition, subjects with abnormal stereopsis demonstrated decreased three-dimensional perception during stereoscopic display viewing.[5] Even microstrabismus with deviations of less than 6 prism diopters may be associated with central suppression and impaired binocular visual function.[6] Therefore, early diagnosis and appropriate management of strabismus are important for achieving optimal binocular visual outcomes.[7]

Currently, manual clinical examination remains the gold standard for assessing ocular motility and ocular alignment.[8,9] However, accurate evaluation of ocular motility using conventional manual examination techniques requires substantial clinical experience and is subject to examiner-dependent and test-retest variability, particularly when quantifying ocular misalignment.[10,11]

Recently, several studies have explored the application of eye-tracking technology and virtual reality (VR)-based systems for the quantitative assessment of ocular alignment and eye movements to facilitate early diagnosis and precise detection of strabismus.[12–14] However, evidence regarding their application in clinical patient populations remains limited.[15]

Therefore, the present study aimed to evaluate the clinical utility of eye tracking technology using a VR-based oculography headset device (neos®, machineMD, Switzerland) by comparing device-based ocular alignment measurements with conventional manual clinical examinations in patients with strabismus.

Methods

Ethics statement

This retrospective study was approved by the Institutional Review Board of the University of California, Irvine (UCI Study ID: STUDY00000931) and was conducted in accordance with the tenets of the Declaration of Helsinki.

Study design and participants

The medical records of consecutive patients who presented for an initial strabismus evaluation at the Gavin Herbert Eye Institute, University of California, Irvine, were retrospectively reviewed. Patients were included if they had undergone neos® examination as an adjunct to routine clinical evaluation of strabismus. Baseline demographic and ophthalmic characteristics, conventional ocular alignment measurements, and previously acquired neos™ measurements were obtained from the medical records.

The exclusion criteria were paralytic or restrictive strabismus; a history of ocular trauma or ocular surgery, including strabismus surgery; ocular conditions other than strabismus that could affect ocular alignment or visual function, including external ocular, optic nerve, and retinal diseases; poor cooperation during ocular examination; and inability to complete the device-based examination.

Clinical evaluation and ocular examination

Demographic variables included patient age and sex. Baseline ophthalmologic examinations included best-corrected visual acuity (BCVA), spherical equivalent (SE) refractive error, intraocular pressure, stereopsis, and ocular deviation. Stereoacuity was assessed using the Titmus stereotest (Stereo Optical Co., Inc., Chicago, IL, USA). Ocular deviation was measured at distance (6 m) and near (1/3 m) fixation by an experienced ophthalmologist (D.W.S.). Total deviation was measured using the alternate prism cover test (APCT) and represented the sum of tropia and phoria. Tropia was measured using the simultaneous prism cover test (SPCT) under binocular viewing conditions and represented the manifest component only. Horizontal and vertical deviations measured by both APCT and SPCT were included in the analysis.

Machine-based ocular alignment evaluation

Machine-based ocular alignment evaluation was performed using the neos™ system (machineMD AG, Bern, Switzerland), a virtual reality headset–based device equipped with 200-Hz infrared eye tracking.[12] Participants were seated in an upright position and instructed to fixate on visual stimuli presented within the headset. Eye position was recorded using the integrated eye tracker during testing. All device-based examinations were performed by a trained examiner (D.L.) using the Ocular Alignment Test Execution mode.

Machine-generated ocular alignment measurements were obtained during distance fixation. Horizontal and vertical total deviations were derived from the “Tropia + Phoria from alternating stimulus” parameter, whereas horizontal and vertical tropia measurements were derived from the “Tropia from binocular stimulus” parameter.

To ensure the reliability of device-based measurements, tracking precision (°), representing the stability and consistency of eye tracking, and data completeness (%), representing the proportion of successfully acquired data during the examination, were reviewed as quality indicators. For data quality control, only measurements with tracking precision ≤ 1.5° and data completeness ≥ 80% in both eyes were included in the analysis.

Comparison of manual and machine-based measurements

Manual and machine-based measurements of horizontal and vertical ocular deviation were compared in all included patients. Examiner-measured total horizontal and vertical deviations obtained using APCT were compared with the corresponding “Tropia + Phoria from alternating stimulus” measurements generated by the neos™ system. Examiner-measured horizontal and vertical tropia obtained using SPCT were compared with the corresponding “Tropia from binocular stimulus” measurements generated by the neos™ system. Measurement differences and agreement between manual and machine-based measurements were additionally evaluated in the two groups.

To further evaluate agreement according to deviation magnitude, subgroup analyses were performed based on the magnitude of total horizontal deviation measured by APCT at distance fixation. Patients were classified into < 15 PD and ≥ 15 PD groups, irrespective of whether the horizontal deviation was exotropic or esotropic. Associations and measurement differences between manual and machine-based measurements were additionally evaluated in the two groups.

Statistical analysis

Statistical analyses were performed using SPSS Statistics version 24.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean ± standard deviation. Paired manual and device-based measurements of horizontal and vertical ocular deviation were compared using the Wilcoxon signed-rank test. Associations between manual and device-based measurements were assessed using Spearman’s rank correlation coefficient (ρ). Agreement between the two measurement methods was evaluated using Bland–Altman analysis, with mean differences and 95% limits of agreement (LoA) calculated for each measurement parameter.

Differences between manual and device-based measurements were calculated as manual minus device-based measurements; therefore, positive values indicated underestimation by the device, whereas negative values indicated overestimation. To investigate whether agreement between manual and device-based measurements varied according to the magnitude of horizontal deviation, subgroup analyses were performed using a 15-PD cutoff. Manual–device measurement differences between the ≥ 15 PD and < 15 PD groups were compared using the Mann–Whitney U test. All statistical tests were two-sided, and a P value < 0.05 was considered statistically significant.

Results

A total of 46 patients were included in the analysis, comprising 21 males and 25 females, with a mean age of 37.61 ± 20.41 years (range, 8–85 years) (Table 1). The mean horizontal total deviation by APCT was 11.37 ± 15.95 PD, and the mean horizontal tropia by SPCT was 8.11 ± 11.84 PD. The mean vertical total deviation by APCT was 2.09 ± 3.84 PD, and the mean vertical tropia by SPCT was 1.54 ± 3.74 PD.

Table 1
Baseline characteristics of the study population according to horizontal deviation magnitude (n = 46).

Characteristic	Total (n = 46)	≥ 15 PD group (n = 16)	< 15 PD group (n = 30)	P-value
Gender (male:female)	21:25	9:7	12:18	0.292*
Age at initial visit (year)	37.61 ± 20.41 (8 to 85)	47.38 ± 25.46	32.40 ± 15.18	0.090†
BCVA (logMAR)				
OD	0.06 ± 0.11 (0.00 to 0.54)	0.06 ± 0.06	0.06 ± 0.13	0.084†
OS	0.05 ± 0.10 (0.00 to 0.40)	0.08 ± 0.11	0.04 ± 0.08	0.147†
SE refractive errors (D)				
OD	-0.27 ± 1.59 (-4.00 to + 2.75)	0.20 ± 1.68	-0.52 ± 1.51	0.055†
OS	-0.28 ± 1.59 (-6.00 to + 2.50)	-0.06 ± 2.03	-0.41 ± 1.33	0.057†
IOP (mmHg)				
OD	14.48 ± 3.47 (8 to 22)	14.75 ± 4.12	14.33 ± 3.13	0.935†
OS	14.50 ± 3.64 (8 to 22)	14.63 ± 4.22	14.43 ± 3.37	0.963†
Titmus test score (arcsec)	574.57 ± 112.14 (40 to 4000)	800.00 ± 133.59	288.00 ± 885.42	< 0.001†
Tracking precision (degree, °)				
OD	0.56 ± 0.34 (0.2 to 1.5)	0.67 ± 0.33	0.50 ± 0.33	0.058
OS	0.59 ± 0.35 (0.1 to 1.5)	0.73 ± 0.37	0.52 ± 0.32	0.053
Data completeness (%)				
OD	91.79 ± 5.50 (80.10 to 99.70)	92.78 ± 5.93	91.26 ± 5.28	0.263
OS	91.19 ± 5.89 (80.20 to 99.40)	92.67 ± 5.29	90.60 ± 5.67	0.209
Manual horizontal deviation at initial visit (PD)				

Characteristic	Total (n = 46)	≥ 15 PD group (n = 16)	< 15 PD group (n = 30)	P-value
Angle of total deviation by APCT	11.37 ± 15.95 (0 to 70)	27.81 ± 17.17	2.60 ± 3.68	< 0.000†
Angle of tropia by SPCT	8.11 ± 11.84 (0 to 50)	20.81 ± 12.03	1.33 ± 2.48	< 0.000†
Manual vertical deviation at initial visit (PD)				
Angle of total deviation by APCT	2.09 ± 3.84 (0 to 50)	3.12 ± 3.34	1.53 ± 4.02	< 0.000†
Angle of tropia by SPCT	1.54 ± 3.74 (0 to 20)	2.00 ± 3.33	1.30 ± 3.98	< 0.000†

Values are presented as mean ± standard deviation (range) unless otherwise indicated.

Significant deviation was defined as a horizontal deviation of ≥ 15 prism diopters measured at the initial visit. Mild deviation was defined as a horizontal deviation of < 15 prism diopters, including orthophoria and small-angle phoria.

BCVA = best corrected visual acuity; logMAR = logarithm of the minimum angle of resolution; OD = oculus dexer; OS = oculus sinister;

SE = spherical equivalent; D = diopter; PD = prism diopter, APCT= Alternating prism and cover test, SPCT = Simultaneous prism and cover test

*Pearson's chi-square test; †Mann-Whitney test.

To evaluate differences according to deviation magnitude, patients were stratified into two subgroups based on horizontal total deviation by APCT at distance fixation: the ≥ 15 PD group (n = 16) and the < 15 PD group (n = 30). Baseline characteristics, including age, BCVA, SE refractive error, and intraocular pressure, did not differ significantly between the two groups. Similarly, the device-based quality indicators, including tracking precision and data completeness, did not differ significantly between the two groups. Additionally, the Titmus stereotest score was significantly higher in the ≥ 15 PD group than in the < 15 PD group ($P < 0.001$), indicating poorer stereopsis in patients with larger horizontal deviations.

We compared manual and device-based measurements across four parameters, categorized by total versus tropia deviation and horizontal versus vertical components (Table 2). No significant differences between manual and device-based measurements were observed for either horizontal total deviation by APCT ($P = 0.940$) or horizontal tropia by SPCT ($P = 0.074$). In contrast, significant differences were observed for both vertical total deviation by APCT and vertical tropia by SPCT, with mean differences of − 1.30 ± 4.96 PD and − 2.22 ± 5.34 PD, respectively, indicating significantly larger device-based than manual measurements for both vertical parameters.

Table 2
Comparison, correlation, and agreement between manual and device-based measurements in the entire cohort (n = 46).

Pairwise comparison	Mean difference ± SD	P-value*	95% limits of agreement
Horizontal total deviation by APCT (PD)			
Manual – Device	-0.30 ± 8.22	0.940	-16.42 to 15.82
Vertical total deviation by APCT (PD)			
Manual – Device	-1.30 ± 4.96	0.020	-11.02 to 8.41
Horizontal tropia by SPCT (PD)			
Manual – Device	-2.30 ± 8.04	0.074	-18.06 to 13.45
Vertical tropia by SPCT (PD)			
Manual – Device	-2.22 ± 5.34	< 0.001	-12.68 to 8.24

SD = standard deviation.

Difference was calculated as manual measurement minus device measurement.

APCT = Alternating prism and cover test; SPCT = Simultaneous prism and cover test

**P*-value = Wilcoxon signed-rank test

Despite the significant correlations, the 95% limits of agreement varied across the four parameters and remained relatively wide, indicating considerable individual-level variability between manual and device-based measurements.

Scatter plot analysis demonstrated significant positive correlations between manual and device-based measurements across all four parameters (Fig. 1). Strong positive correlations were observed for horizontal total deviation by APCT (Spearman ρ = 0.780, $P < 0.001$), vertical total deviation by APCT (ρ = 0.728, $P < 0.001$), and horizontal tropia by SPCT (ρ = 0.783, $P < 0.001$). In contrast, vertical tropia by SPCT showed a weaker, although statistically significant, positive correlation (ρ = 0.424, $P = 0.003$).

Bland–Altman analysis demonstrated varying levels of agreement between manual and device-based measurements across the four parameters (Fig. 2). Horizontal total deviation by APCT showed minimal systematic bias, with a mean difference of − 0.30 PD, whereas the 95% limits of agreement were relatively wide, ranging from − 16.42 to 15.81 PD (Fig. 2A). Similar patterns of relatively small mean differences but wide limits of agreement were observed for the other three parameters (Fig. 2B–D), indicating considerable individual-level variability between the two measurement methods. Notably, visual inspection of the Bland–Altman plots showed a tendency toward greater dispersion of manual–device differences as the magnitude of ocular deviation increased across all four parameters.

Table 3 shows the comparison of manual–device measurement differences and variability according to horizontal deviation magnitude. No significant between-group differences in manual–device measurement differences were observed between patients with larger horizontal deviations of ≥ 15 PD and those with smaller deviations of < 15 PD for horizontal total deviation by APCT (*P* = 0.430), vertical total deviation by APCT (*P* = 0.121), or horizontal tropia by SPCT (*P* = 0.288), suggesting that the magnitude of systematic discrepancy between manual and device-based measurements was generally comparable across small- and large-angle deviations for these parameters. In contrast, vertical tropia by SPCT showed a significant between-group difference in measurement difference (P = 0.010).

Table 3
Comparison of measurement differences and limits of agreement between groups with horizontal deviation ≥ 15 PD and < 15 PD.

Measurement difference	**≥ 15 PD group (n = 16)**		**< 15 PD group (n = 30)**		**Between-group comparison**	
(Manual − Device, PD)	**Median (IQR)**	**95% LoA**	**Median (IQR)**	**95% LoA**	**Difference P-value***	**Variability P-value†**
Horizontal total deviation by APCT	0.00 (−1.75 to 7.00)	−24.49 to 24.99	0.00 (−2.00 to 2.00)	−9.86 to 8.66	0.430	0.016
Vertical total deviation by APCT	−1.00 (−3.75 to 0.00)	−16.91 to 10.91	0.00 (−1.00 to 0.00)	−6.48 to 5.68	0.121	0.003
Horizontal tropia by SPCT	−2.00 (−9.50 to 1.75)	−29.10 to 18.60	0.00 (−1.25 to 0.00)	−8.66 to 7.19	0.288	< 0.001
Vertical tropia by SPCT	−1.50 (−7.50 to −1.00)	−18.71 to 8.84	−1.00 (−1.00 to 0.00)	−7.67 to 6.13	0.010	0.003

Values represent the difference between manual and device measurements (manual − device) and are presented as median (interquartile range [IQR]).

The 95% limits of agreement (LoA) were calculated as the mean difference ± 1.96 SD.

Total deviation was measured using the APCT and represents the sum of tropia and phoria.

Tropia deviation was measured using the SPCT under binocular viewing conditions and represents the tropic component only.

APCT = alternating prism cover test; SPCT = simultaneous prism cover test; PD = prism diopters; IQR = interquartile range.

*P-value for between-group comparison of measurement differences using the Mann–Whitney U test.

† P-value for comparison of variance using Levene's test.

Additionally, the variability of manual–device measurement differences was compared between the two groups using Levene's test. Measurement variability was significantly greater in the ≥ 15 PD group across all four parameters, as reflected by the wider 95% limits of agreement. These findings indicate that although larger horizontal deviations did not substantially increase systematic manual–device discrepancies for most parameters, individual-level agreement became less consistent in the ≥ 15 PD group.

To visually illustrate the differences in measurement variability according to horizontal deviation magnitude, the 95% limits of agreement for the ≥ 15 PD and < 15 PD groups were graphically compared (Fig. 3). The ≥ 15 PD group consistently showed wider 95% limits of agreement across all four parameters, corresponding to significantly greater measurement variability than the < 15 PD group. Levene's test confirmed significant differences in variability for horizontal total deviation by APCT (P = 0.016), vertical total deviation by APCT (P = 0.003), horizontal tropia by SPCT (P < 0.001), and vertical tropia by SPCT (P = 0.003).

Discussion

In this study, VR-based ocular alignment measurements showed significant positive correlations with conventional manual measurements across all four evaluated parameters. Particularly strong correlations were observed for horizontal total deviation by APCT and horizontal tropia by SPCT. No significant differences between manual and device-based measurements were observed for the two horizontal parameters, whereas device-based measurements were significantly larger than manual measurements for both vertical parameters. Bland–Altman analysis further demonstrated relatively wide limits of agreement, indicating considerable individual-level variability despite the overall correlations between the two methods. Importantly, subgroup analysis showed significantly greater measurement variability in patients with horizontal deviations ≥ 15 PD across all four parameters, suggesting reduced consistency of device-based measurements at larger deviation magnitudes.

Accurate detection and quantitative assessment of ocular misalignment are important for the diagnosis and management of strabismus, particularly in children at risk of amblyopia and impaired binocular visual development.[16] However, small-angle strabismus, including microstrabismus, may be difficult to detect during routine clinical examination, particularly in nonspecialized settings.[17,18] Delayed diagnosis may lead to prolonged treatment and poorer binocular visual outcomes.[4,19] Furthermore, strabismus care has become increasingly subspecialized, with management often concentrated among a limited number of experienced specialists.[20]

In recent years, several attempts have been made to utilize eye-tracking technology for automated assessment of ocular alignment and eye movements to facilitate early diagnosis and precise detection of strabismus.[12–14,21] However, considerable heterogeneity exists in device architecture, testing protocols, and validation methods,[22,23] and evidence regarding their performance in clinical populations with strabismus remains limited.[12] In the present study, we assessed the clinical utility of the neos™

system by comparing device-based ocular alignment measurements with conventional manual examinations in patients with strabismus.

Several studies have demonstrated the feasibility of VR-based technologies for automated strabismus screening and diagnosis. Wang et al.[22] reported that a VR-based strabismus screening system showed a sensitivity of 80.3% and a specificity of 80%, with moderate agreement with manual examinations. Similarly, Yeh et al.[23] demonstrated good agreement between VR-based ocular deviation measurements and APCT findings, particularly during near fixation testing.

In the present study, no significant differences were observed between manual and device-based measurements for either horizontal parameter in Table 2, indicating generally comparable measurements for horizontal deviations. In contrast, device-based measurements were significantly larger than manual measurements for both vertical parameters, by an average of 1.30 PD for vertical total deviation by APCT and 2.22 PD for vertical tropia by SPCT. These findings suggest that measurement performance may differ between horizontal and vertical ocular deviations.

Additionally, correlation analysis demonstrated significant positive associations between manual and device-based measurements across all four parameters in Fig. 1, further supporting the potential value of VR- and AI-based technologies for strabismus assessment.[24] Among these parameters, the strongest correlations were observed for horizontal total deviation by APCT and horizontal tropia by SPCT, with Spearman ρ values of 0.780 and 0.783, respectively. In contrast, vertical tropia by SPCT showed the weakest correlation, with a Spearman ρ of 0.424. Although the vertical parameters remained significantly correlated with manual measurements, device-based measurements tended to be larger than manual measurements, indicating the need for cautious interpretation of vertical deviation measurements.

Despite these correlations, Bland–Altman analysis revealed relatively wide 95% limits of agreement across all four parameters. In particular, horizontal total deviation by APCT showed minimal systematic bias, with a mean difference of − 0.30 PD, whereas the 95% limits of agreement ranged from − 16.42 to 15.81 PD. Therefore, substantial differences between manual and device-based measurements may occur at the individual-patient level despite the overall correlation between the two methods.

Furthermore, in the comparison of manual–device measurement differences according to horizontal deviation magnitude, no significant differences were observed between the ≥ 15 PD and < 15 PD groups for three of the four parameters. This suggests that manual and device-based measurements were comparable for these parameters regardless of horizontal deviation magnitude. Vertical tropia by SPCT was the only parameter showing a significant between-group difference, with a greater discrepancy between manual and device-based measurements in the ≥ 15 PD group. This finding is consistent with the preceding analysis showing larger device-based measurements for vertical deviations, suggesting the need for careful interpretation of vertical measurements, particularly in patients with larger horizontal deviations.

In the comparison of measurement variability between the two groups, the ≥ 15 PD group showed significantly greater variability across all four parameters, with wider 95% limits of agreement than the < 15 PD group. These findings indicate lower individual-level measurement consistency in patients with larger horizontal deviations, suggesting that device-based measurements in large-angle strabismus should be interpreted in conjunction with a detailed clinical assessment of ocular alignment.

The greater measurement variability observed in patients with larger deviations can be explained by several factors. First, horizontal deviation in patients with large-angle strabismus can fluctuate depending on the patient's condition.[25–28] Hatt et al.[28] reported that the control of intermittent exotropia varies throughout the day and even within minutes, indicating that a single measurement may not fully represent the severity of deviation. In addition, Economides et al.[27] reported that ocular deviation in patients with strabismus exhibits inherent biological variability. Therefore, part of the discrepancy observed between manual and device-based measurements may reflect the inherent biological variability of strabismus rather than device-related measurement error alone. Second, manual clinical examination allows the examiner to assess the patient's level of control and adjust examination conditions accordingly. During APCT, the duration of dissociation and repeated measurements can be adjusted according to the clinical response. Economides et al.[27] described that APCT involves repeated alternate occlusion and adjustment of prism strength until no refixation movement is observed, allowing the examiner to determine ocular deviation based on repeated observations. In contrast, device-based examinations are performed using a standardized testing sequence and may not fully account for individual variations in fusional control or fixation behavior. These differences in examination conditions may have contributed to the discrepancy observed between manual and device-based measurements. Nevertheless, the overall correspondence between device-based and manual measurements suggests that VR-based oculography may be useful for quantitative ocular alignment assessment, particularly in patients with smaller deviations.

There are several limitations to this study. The sample size was relatively small, which may have limited the statistical power of subgroup analyses; therefore, separate analyses according to strabismus type and level of control were not performed. Patients with exotropia and esotropia were analyzed together, although ocular alignment characteristics and fusional control may differ between these groups. In addition, the potential influence of strabismus control on the variability of ocular deviation measurements could not be separately evaluated .[28] Lastly, only distance deviation measurements were analyzed because the VR-based system did not provide near deviation measurements. Assessment of ocular alignment at both distance and near is clinically important because the distance–near relationship contributes to the classification and management of strabismus.[16,29] Future studies with larger sample sizes and standardized control grading systems are needed to further evaluate device performance according to strabismus subtype and control status.

In conclusion, VR-based oculography demonstrated significant correlations with conventional manual measurements of ocular alignment, particularly for horizontal deviations. However, individual-level

variability increased significantly in patients with larger horizontal deviations. These findings support the potential utility of VR-based oculography as an objective quantitative tool for ocular alignment assessment, while highlighting the need to interpret individual measurements cautiously, particularly in large-angle strabismus. Further studies with larger cohorts and stratification by strabismus subtype and control status are warranted to refine its clinical applicability.

Declarations

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing Interests:

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval:

This retrospective study was approved by the Institutional Review Board of the University of California, Irvine (UCI Study ID: STUDY00000931) and was conducted in accordance with the tenets of the Declaration of Helsinki.

Informed Consent:

The requirement for informed consent was waived by the Institutional Review Board because of the retrospective nature of the study.

Funding:

No funding was received for this work

Author Contribution

D.L. conceived and designed the study, collected and analyzed the data, and drafted the manuscript. K.R. contributed to IRB preparation and regulatory documentation and edited and revised the manuscript. M.J.W. contributed to data collection and analysis. D.W.S. supervised the study, contributed to study design and interpretation of the results, and critically revised the manuscript. All authors reviewed and approved the final manuscript.

Data Availability:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Fieß A, Gißler S, Hartmann A et al (2026) Prevalence of strabismus, nystagmus and risk factors in children and adolescents born preterm with and without retinopathy of prematurity: results from the Gutenberg Prematurity Study Young. Br J Ophthalmol 110(4):370–378
2. Fieß A, Kölb-Keerl R, Schuster AK et al (2017) Prevalence and associated factors of strabismus in former preterm and full-term infants between 4 and 10 Years of age. BMC ophthalmol 17(1):228
3. Hashemi H, Pakzad R, Heydarian S et al (2019) Global and regional prevalence of strabismus: a comprehensive systematic review and meta-analysis. Strabismus 27(2):54–65
4. Wiecek E, Kosovicheva A, Ahmed Z et al (2024) Peripheral binocular imbalance in anisometropic and strabismic amblyopia. Invest Ophthalmol Vis Sci 65(4):36–36
5. Kim S, Suh Y, Yun C, Yoo E, Yeom J, Cho Y (2013) Influence of stereopsis and abnormal binocular vision on ocular and systemic discomfort while watching 3D television. Eye(Lond) 27(11):1243–1248
6. Chima AS, Formankiewicz MA, Waugh SJ (2016) Interocular suppression patterns in binocularly abnormal observers using luminance-and contrast-modulated noise stimuli. J Vis 16(10):20–20
7. Garcia CAA, ABd S, LLd MendonçaMBMA, Oréfice F (2004) Prevalence of strabismus among students in Natal/RN-Brazil. Arq Bras Oftalmol 67(5):791–794
8. Thurtell MJ, Leigh RJ (2011) Nystagmus and saccadic intrusions. Handb Clin Neurol 102:333–378
9. Rucker JC, Kennard C, Leigh RJ (2011) The neuro-ophthalmological examination. Handb Clin Neurol 102:71–94
10. Stunkel L, Newman-Toker DE, Newman NJ, Biousse V (2021) Diagnostic error of neuro-ophthalmologic conditions: state of the science. J Neuroophthalmol 41(1):98–113
11. Holmes JM, Leske DA, Hohberger GG (2008) Defining real change in prism-cover test measurements. Am J Ophthalmol 145(2):381–385
12. Coito A, Naidu A, Lehmann J, Hauser B, Brügger D, Abegg M (2025) Test-retest reliability of gaze precision of a novel virtual reality-based medical device. Front Virtual Real 6:1502679
13. González-Vides L, Hernández-Verdejo JL, Cañadas-Suárez P (2023) Eye tracking in optometry: A systematic review. J Eye Mov Res 16(3):15
14. Rizzo J-R, Beheshti M, Dai W, Rucker JC (2019) Eye movement recordings: Practical applications in neurology. Semin Neurol 39(6):775–784
15. Reuter E, Luerweg J, Schneider R et al (2025) Case Report: Improvement of diplopia due to severe internuclear ophthalmoplegia by 4-aminopyridine documented using a novel virtual reality-based oculography headset. Front Virtual Real 6:1595694

16. Sprunger DT, Lambert SR, Hercinovic A et al (2023) Esotropia and exotropia preferred practice pattern®. Ophthalmology 130(3):P179–P221
17. Rucker JC (2019) Nystagmus and saccadic intrusions. Continuum (Minneap Minn) 25(5):1376–1400
18. Helveston EM, von Noorden GK (1967) Microtropia: a newly defined entity. Arch Ophthalmol 78(3):272–281
19. Taylor K, Elliott S (2014) Interventions for strabismic amblyopia. Cochrane Database Syst Rev 2014(7):CD006461
20. Campbell RJ, Gill SS, ten Hove M et al (2015) Strabismus surgical subspecialization: a population-based analysis. JAMA Ophthalmol 133(5):555–559
21. Pueyo V, Pérez-Roche T, Prieto E et al (2020) Development of a system based on artificial intelligence to identify visual problems in children: study protocol of the TrackAI project. BMJ open 10(2):e033139
22. Wang Y-m, Liu J, Chen W-w, Jiang M-x, Fu J (2024) Accuracy of virtual reality in automated measurement and diagnosis of strabismus. Digit Health 10:20552076241308713
23. Yeh P-H, Liu C-H, Sun M-H, Chi S-C, Hwang Y-S (2021) To measure the amount of ocular deviation in strabismus patients with an eye-tracking virtual reality headset. BMC ophthalmol 21(1):246
24. Wu D, Li Y, Zhang H et al (2024) An artificial intelligence platform for the screening and managing of strabismus. Eye (Lond) 38(16):3101–3107
25. Guo Y, Guan Y, Li L, Jiang J (2025) Multiple Factors Causing Variability of Alignment in Childhood Concomitant Strabismus. Am J Ophthalmol 270:77–82
26. Kim WJ, Kim MM (2017) Variability of preoperative measurements in intermittent exotropia and its effect on surgical outcome. J AAPOS 21(3):210–214
27. Economides JR, Adams DL, Horton JC (2016) Variability of ocular deviation in strabismus. JAMA ophthalmol 134(1):63–69
28. Hatt SR, Mohney BG, Leske DA, Holmes JM (2008) Variability of control in intermittent exotropia. Ophthalmology 115(2):371–376 e2
29. Burke J (2015) Distance–near disparity esotropia: can we shrink the gap? Eye(Lond) 29(2):208–213

Figures

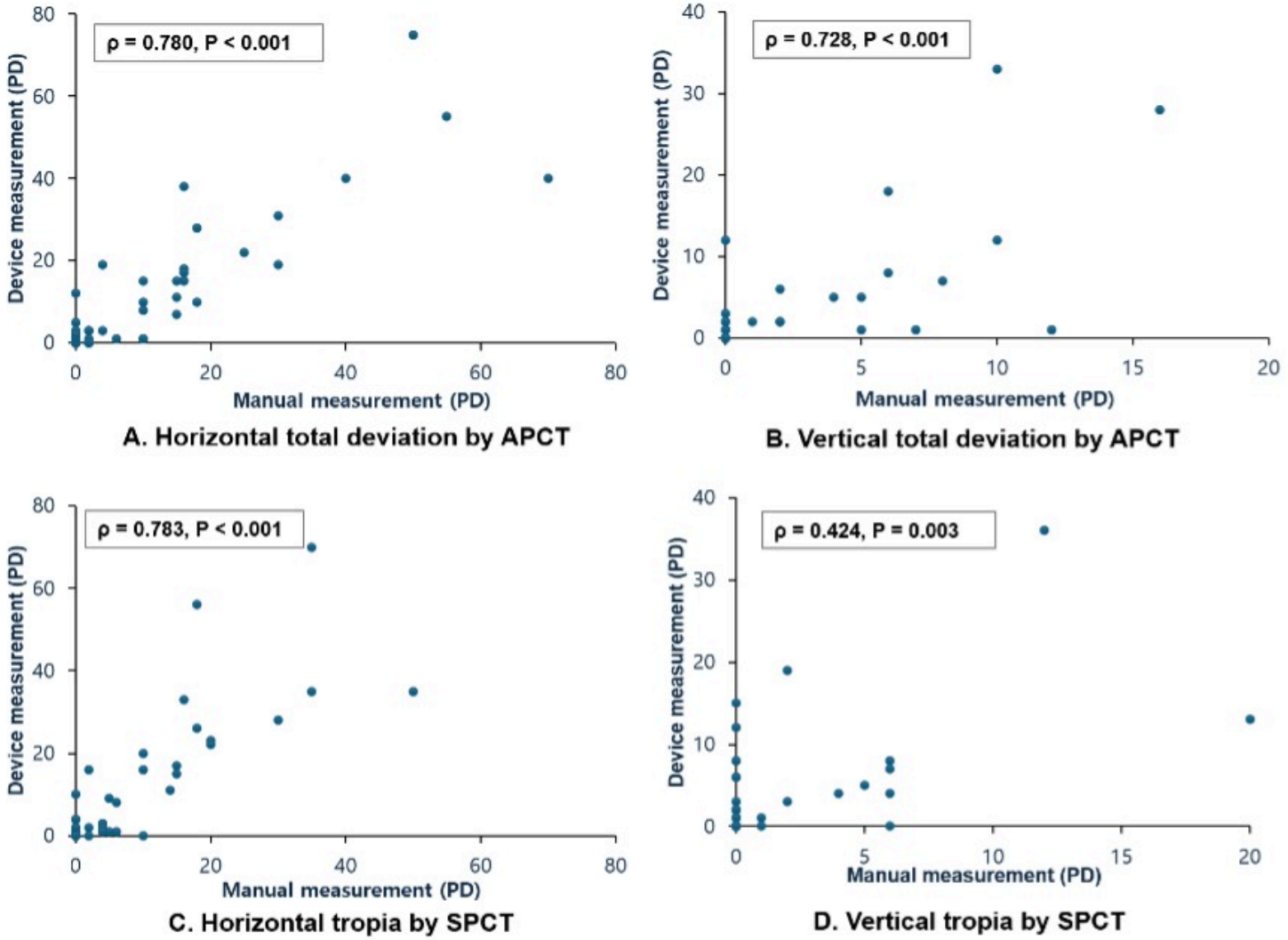


Figure 1

Scatter plots illustrating the relationship between manual and device-based measurements. (A) Horizontal total deviation by APCT, (B) Vertical total deviation by APCT, (C) Horizontal tropia by SPCT, and (D) Vertical tropia by SPCT. Strong positive correlations were observed for horizontal total deviation (Spearman ρ = 0.780, P < 0.001), vertical total deviation (ρ = 0.728, P < 0.001), and horizontal tropia (ρ = 0.783, P < 0.001). In contrast, vertical tropia showed a weaker, although statistically significant, positive correlation (ρ = 0.424, P = 0.003). Each point represents an individual subject. APCT = alternating prism cover test; SPCT = simultaneous prism cover test; PD = prism diopters.

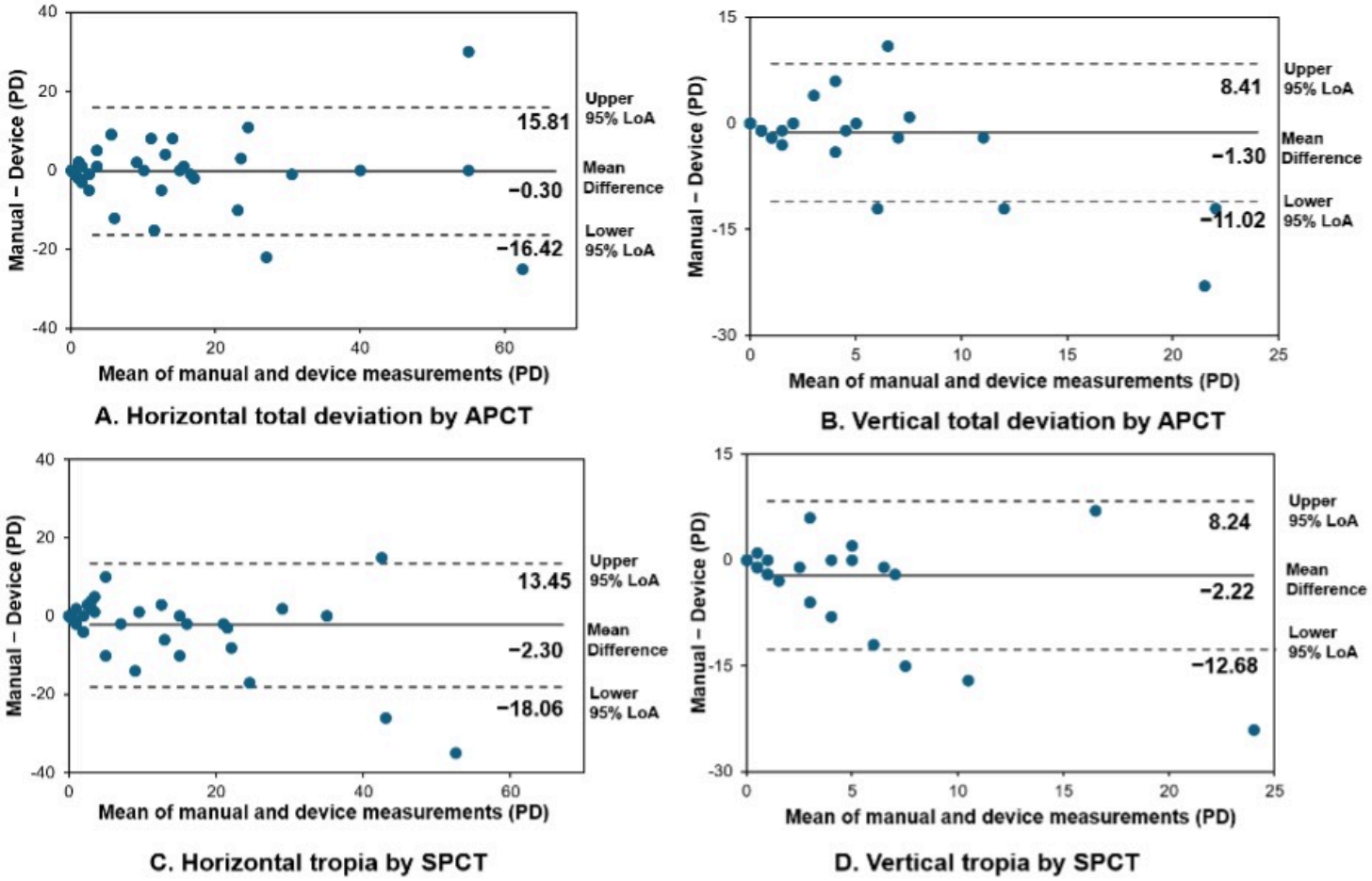


Figure 2

Bland–Altman analysis of agreement between manual and device-based measurements. (A) Horizontal total deviation by APCT, (B) vertical total deviation by APCT, (C) horizontal tropia by SPCT, and (D) vertical tropia by SPCT. Differences were calculated as manual minus device-based measurements The mean differences (manual − device) ranged from −0.30 to −2.30 PD across the four parameters, with the smallest bias observed for horizontal total deviation by APCT. The 95% limits of agreement were wide, indicating considerable individual-level variability between the two measurement methods. . The relatively wide limits of agreement, particularly for horizontal tropia, indicate substantial variability at the individual-subject level. The solid line represents the mean difference, and the dashed lines represent the 95% limits of agreement. APCT = alternating prism cover test; SPCT = simultaneous prism cover test; PD = prism diopters; LoA = limits of agreement.

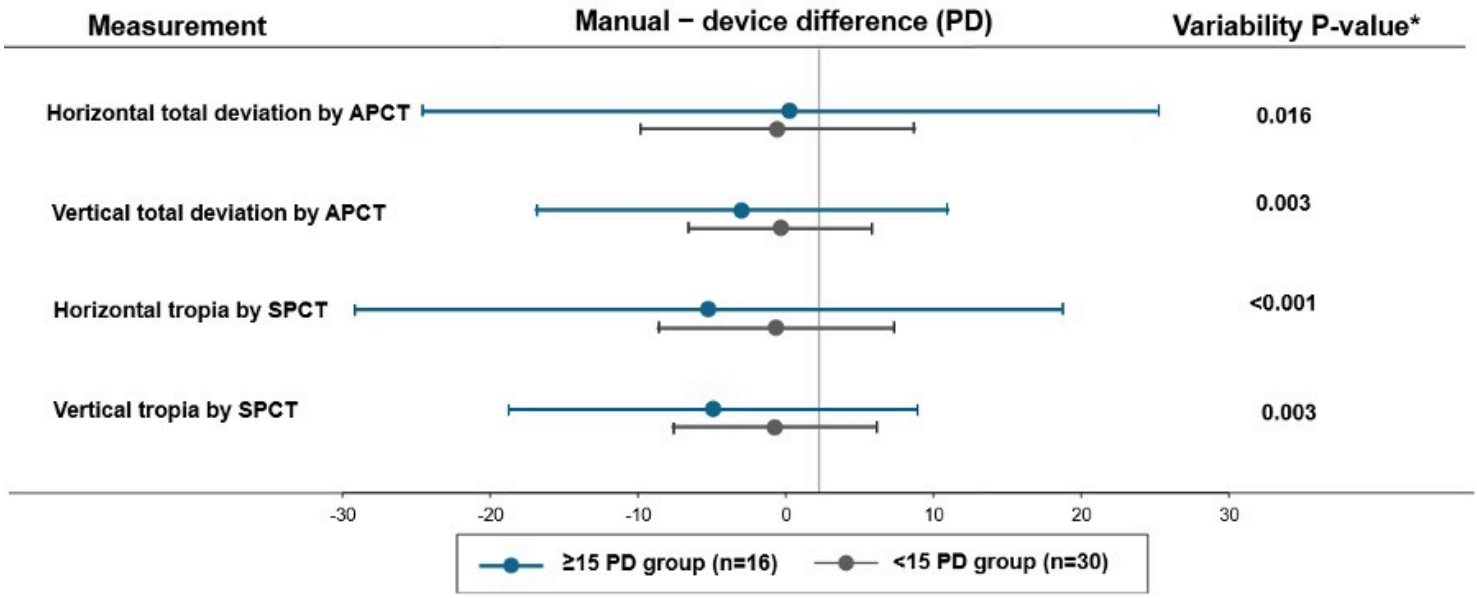


Figure 3

Graphical comparison of 95% limits of agreement according to horizontal deviation magnitude. Points indicate mean differences (manual − device), and horizontal lines represent the 95% limits of agreement (LoA). The ≥15 PD group showed significantly greater measurement variability and wider LoA than the <15 PD group across all four measurements. *P values for variability were determined using Levene's test. APCT = alternating prism cover test; SPCT = simultaneous prism cover test; PD = prism diopters; LoA = limits of agreement.