
Conventional Versus Virtual Reality-Based Hess-Lancaster Assessment: Agreement and Repeatability in Ocular Motility Evaluation

[Francisco Javier Povedano-Montero](#)*, [Alvaro Perales-Serrano](#), [Daniela León Lobo](#), [Rut González-Jiménez](#), [Ricardo Bernárdez-Vilaboá](#), [Juan E. Cedrún-Sánchez](#)

Posted Date: 25 May 2026

doi: 10.20944/preprints202605.1592.v1

Keywords: Hess-Lancaster test; virtual reality; Dicopt Pro; ocular motility; strabismus; Bland-Altman analysis; repeatability



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Conventional Versus Virtual Reality-Based Hess-Lancaster Assessment: Agreement and Repeatability in Ocular Motility Evaluation

Francisco Javier Povedano-Montero ^{1,2,*}, Alvaro Perales-Serrano ¹, Daniela León Lobo ¹, Rut González-Jiménez ¹, Ricardo Bernárdez-Vilaboa ¹ and Juan E. Cedrún-Sánchez ^{1,3}

¹ Optometry and Vision Department, Faculty of Optics and Optometry, Complutense University of Madrid, 28037 Madrid, Spain

² Hospital Doce de Octubre Research Institute (i+12), 28041 Madrid, Spain

³ Applied Vision Research Group, Faculty of Optics and Optometry, Universidad Complutense de Madrid, 28037 Madrid, Spain

* Correspondence: franpove@ucm.es

Abstract

Background: The conventional Hess-Lancaster test is widely used to assess ocular misalignment across diagnostic gaze positions, but it relies on subjective responses and manual recording. Virtual reality may provide a more standardized framework for ocular motility assessment. **Objectives:** To evaluate the agreement and within-method repeatability of point-by-point deviation measurements obtained with the conventional Hess-Lancaster test and a VR-based Hess-Lancaster assessment implemented in Dicopt Pro. **Methods:** This cross-sectional observational study included 52 adults with suspected or diagnosed ocular motility disorders. Participants underwent both assessments using the same predefined gaze positions. Agreement was assessed using Bland–Altman analysis, concordance correlation coefficients, mean absolute differences, and mixed-effects modeling. Repeatability was evaluated in a subset with repeated measurements using session-to-session differences and intraclass correlation coefficients. **Results:** The VR-based assessment showed moderate agreement with the conventional test, with a mean concordance correlation coefficient of 0.57 for both eyes. Mean bias was 1.22 prism diopters for the right eye and 0.10 prism diopters for the left eye. Repeatability was moderate-to-good, with ICC values ranging from 0.62 to 0.83. **Conclusions:** The VR-based Hess-Lancaster assessment showed small mean differences and moderate agreement with the conventional test, although both methods should be interpreted within the context of the complete clinical examination.

Keywords: Hess-Lancaster test; virtual reality; Dicopt Pro; ocular motility; strabismus; Bland–Altman analysis; repeatability

1. Introduction

Accurate quantification of ocular misalignment across different gaze positions is fundamental in the clinical evaluation of incomitant strabismus and paralytic ocular motility disorders [1,2]. Although prism cover testing provides an angular estimate of deviation in primary position, it does not characterize gaze-dependent changes that are often essential for diagnosing muscle paresis patterns and planning management [1]. For this reason, screen-based mapping techniques such as the Hess test have historically been used to represent binocular misalignment across the classical nine-position gaze framework, allowing systematic assessment of gaze-dependent ocular deviation patterns [2,3].

Since its original description in the early twentieth century, the Hess screen has remained a cornerstone in orthoptic and neuro-ophthalmological practice [2,4]. By dissociating binocular vision

through color separation and plotting perceived alignment errors, the test enables inference of muscle underaction, secondary contracture, and incomitance patterns. However, the conventional procedure depends on subjective patient responses and manual plotting by the examiner, introducing potential sources of variability related to sensory adaptation, fixation instability, and graphical interpretation [3].

Methodological comparisons between dissociative screen tests have demonstrated that measurement outcomes may vary according to test geometry and viewing distance. In particular, divergence-related bias has been described when comparing Hess and Harms screen techniques, suggesting that recorded deviation magnitude can be influenced by experimental configuration [5]. Such findings underscore that screen-based measurements are not purely intrinsic ocular values but partially dependent on test design.

Efforts to increase objectivity in ocular motility assessment have evolved over several decades. Three-dimensional adaptations of the Hess paradigm employing scleral search coils demonstrated the feasibility of recording binocular eye rotations with high spatial precision, including torsional components [6]. These investigations provided valuable insight into ocular kinematics but remained limited to research settings due to their invasive nature and technical complexity. More clinically accessible approaches subsequently emerged, including head-mounted digital video systems capable of quantifying strabismus deviations in portable formats [7]. Additionally, integration of eye-tracking into Hess or Lancaster-type paradigms has shown that subjective alignment tasks can be transformed into coordinate-based recordings, allowing direct capture of gaze position during structured motility testing [8].

Parallel to these developments, immersive virtual reality (VR) platforms incorporating eye-tracking technology have been increasingly applied in visual science and neuro-visual rehabilitation [9,10]. VR systems provide standardized stimulus presentation, controlled luminance, and stable head positioning within a closed visual environment. In clinical applications, VR devices integrating eye-tracking have enabled real-time fixation monitoring, automatic calibration, and structured data export for quantitative analysis [11]. Such characteristics suggest that immersive platforms may offer advantages in standardization and reproducibility when adapted for oculomotor mapping tasks. Recent advances in digital ophthalmology further emphasize the role of integrated eye-tracking systems as scalable tools for objective clinical measurement and data-driven analysis [12].

Importantly, modernization of measurement techniques should preserve the diagnostic architecture that gives the Hess test its clinical value. The classical nine-position framework provides a symmetrical representation of gaze-dependent deviations and remains central to clinical interpretation of incomitant strabismus [2,3]. Several previous objective approaches modified task geometry or reduced positional sampling, thereby limiting direct comparability with conventional Hess charts [6]. Demonstrating agreement between a traditional Hess screen and an immersive coordinate-based method that maintains nine identical diagnostic positions therefore represents a necessary methodological step.

Beyond simple replication of conventional measurement, objective coordinate capture produces structured datasets that can be analyzed quantitatively. Vector-based representation of deviation across gaze positions enables systematic characterization of incomitance and may facilitate algorithmic pattern recognition in future research [13,14]. Emerging applications of artificial intelligence in ophthalmology increasingly rely on structured eye-tracking datasets for automated classification and decision support [15]. Recent advances in machine learning applied to eye movement analysis have demonstrated the feasibility of data-driven gaze pattern classification and objective biomarker extraction [16]. Insights from medical imaging applications further reinforce the broad applicability of ML methods to structured eye-tracking data in clinical contexts [17]. The increasing availability of standardized eye-tracking data streams aligns with broader developments in computational analysis of oculomotor behavior and AI-assisted interpretation frameworks.

In addition, the growing emphasis on measurement reproducibility and digital standardization in ophthalmology has highlighted the limitations of manually plotted diagnostic charts. Digital eye-

tracking platforms generate high-resolution gaze datasets with temporal sampling that exceeds the capabilities of conventional screen-based mapping. Recent work in computational eye-movement analysis has demonstrated that structured gaze vectors can be modeled quantitatively, enabling objective comparison across sessions and observers and facilitating integration into data-driven clinical workflows [16,17].

However, despite these technological advances, there is still limited evidence directly comparing conventional Hess measurements with immersive virtual reality systems under equivalent testing conditions. In particular, the lack of methodological standardization across studies has hindered the assessment of true agreement and clinical interchangeability between these approaches. From a methodological perspective, validation of emerging digital diagnostic tools requires formal agreement analysis against established clinical standards to evaluate potential interchangeability and reproducibility [18].

The present study compares measurements obtained with the conventional Hess screen and those recorded using an immersive virtual reality system with digital recording capability (Dicopt Pro), preserving identical predefined gaze positions in both modalities.

Importantly, both systems were configured to ensure equivalence in measurement scale, allowing direct comparison without the need for conversion factors. In addition, repeated measurements were performed within each method, enabling not only agreement analysis but also evaluation of within-method repeatability.

The aim of this study was to evaluate the agreement between the conventional Hess-Lancaster test and a virtual reality-based Hess-Lancaster assessment using the same nine-position framework. A secondary objective was to assess within-method repeatability for both procedures. We hypothesized that the virtual reality-based assessment would show limited systematic bias and moderate-to-good repeatability, while recognizing that agreement may vary across gaze positions.

2. Materials and Methods

2.1. Study Design and Participants

A cross-sectional observational study was conducted between October 2025 and February 2026 at the Ophthalmology Department of Hospital Clínico San Carlos (Madrid, Spain). Participants were consecutively recruited from the Pediatric Ophthalmology and Strabismus Unit.

A total of 54 participants were initially enrolled. After applying exclusion criteria, 2 participants were excluded, resulting in a final sample of 52 participants included in the analysis.

Inclusion criteria comprised:

- age ≥ 18 years,
- clinical diagnosis or suspicion of oculomotor disorder (e.g., strabismus, cranial nerve palsy, or mechanical restriction)
- best-corrected visual acuity of at least 0.5 (decimal) in each eye, and
- sufficient cognitive ability to understand and perform the tests.

Exclusion criteria included severe ocular pathology interfering with fixation or binocular vision; ocular surgery within the previous 6 months; neurological conditions affecting ocular motility; and inability to tolerate or properly use the VR headset.

The study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Ethics Committee of Hospital Clínico San Carlos (approval code: 25/147-E, March 6, 2025). Written informed consent was obtained from all participants.

The study workflow, including participant selection, randomization, and assessment procedures, is summarized in Figure 1.

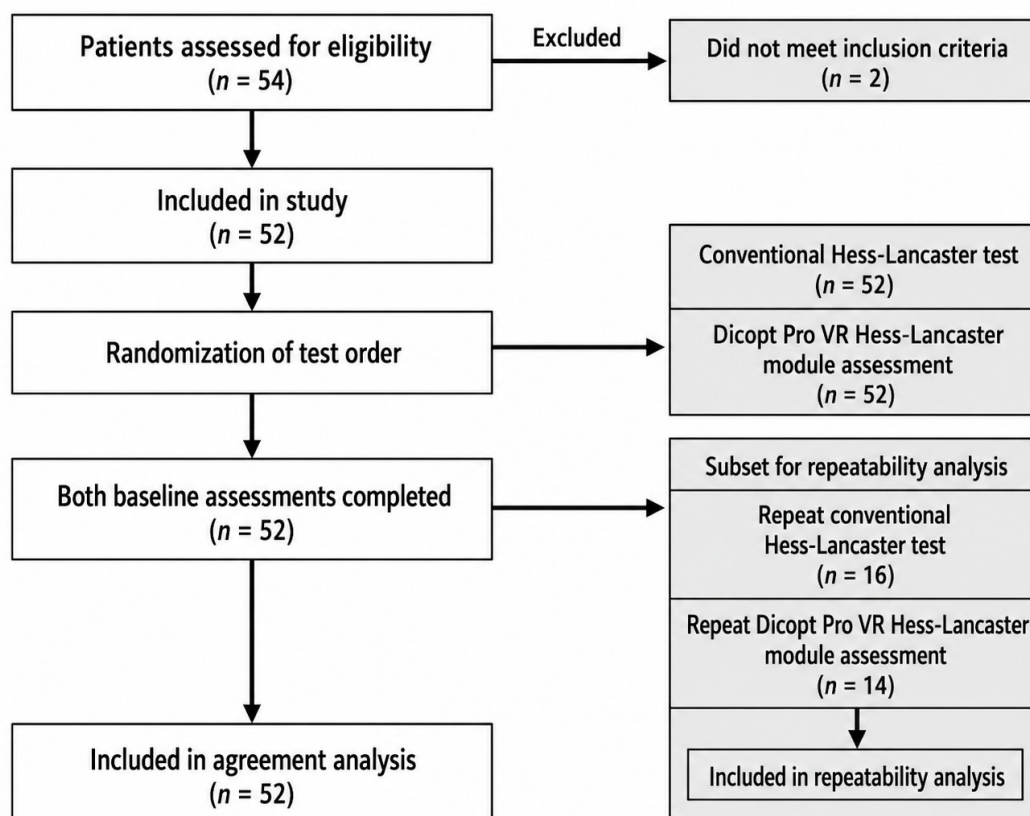


Figure 1. Study flow diagram.

2.2. Sample Size Considerations

Sample size estimation was based on the comparison of paired measurements between two methods. Assuming a conservative standard deviation of the differences ($\sigma = 1.5$) and a minimum clinically relevant difference equivalent to one grid unit, with a confidence level of 95% and statistical power of 80%, the minimum required sample size was estimated at 36 participants.

The final sample of 52 participants exceeded this requirement, ensuring adequate statistical precision for agreement analysis.

2.3. Instrumentation

Conventional Hess-Lancaster Test

The traditional Hess-Lancaster screen consisted of a red grid (108 × 108 cm) with 36 × 36 squares (3 cm each). The assessment was performed at a viewing distance of 60 cm, resulting in each grid unit corresponding to approximately 5 prism diopters (Δ). Binocular dissociation was achieved using red-green filters, and measurements were obtained using a red laser pointer stimulus, allowing evaluation of ocular deviation across the Hess-Lancaster gaze positions.

Virtual Reality System (Dicopt Pro)

The virtual assessment was performed using the Pico Neo 3 Pro Eye headset (Pico Interactive Inc., China) together with Dicopt Pro software (V-Vision, Madrid, Spain). Dicopt Pro includes a virtual reality-based Hess-Lancaster module that reproduces the test in an immersive environment using dichoptic stimulus presentation.

The virtual test was configured at a simulated distance of 90 cm, with each grid unit corresponding to approximately 5 prism diopters. Stimuli were presented sequentially across the diagnostic gaze positions, and responses were recorded automatically in digital format.

Although the device incorporates eye-tracking technology, it was not used for measurement acquisition in this protocol, and responses remained subject-dependent.

The virtual reality-based Hess-Lancaster assessment setup is shown in Figure 2. During the test, participants wore the head-mounted display and used a handheld controller to align the virtual pointer with the sequentially presented stimuli. The examiner monitored the procedure through the connected computer interface, where the coordinate-based responses were automatically recorded in digital format.



Figure 2. Virtual reality-based Hess-Lancaster assessment using Dicopt Pro.

2.4. Scale Standardization Between Methods

Importantly, both systems were configured to ensure equivalence in the measurement scale, with each grid unit corresponding to 5 prism diopters in both modalities. This standardization minimizes geometric scaling bias and allows direct comparison between methods without the need for post hoc conversion factors.

2.5. Experimental Procedure

All participants underwent both assessment methods, the conventional Hess-Lancaster test and the Dicopt Pro virtual reality assessment, within a single session. Agreement analysis between methods was performed using the full sample of 52 participants. Within-method repeatability was assessed in a subset of participants who completed repeated measurements during the same session. Repeated conventional Hess-Lancaster measurements were available for 16 participants, whereas repeated Dicopt Pro measurements were available for 14 participants after excluding incomplete repeated recordings. This subset was used specifically to evaluate session-to-session consistency within each method.

To minimize potential order effects, including learning, fatigue, or adaptation, the sequence of the two assessment methods (conventional Hess-Lancaster and virtual reality test) was randomized

across participants. This randomization ensured that neither method systematically benefited from being performed first.

While the order of the two methods was randomized, the sequence of stimulus presentation within each test followed a predefined standardized order.

For the conventional test, participants were positioned at 60 cm from the screen with head stabilization (chin and forehead support). After binocular dissociation using red–green filters, the examiner projected a reference stimulus while the participant aligned a luminous target across the nine diagnostic gaze positions. Measurements were recorded manually on a standardized grid.

Point stimuli were used to assess ocular deviation across the nine diagnostic gaze positions. The procedure was repeated after reversing the filters to evaluate both eyes.

For the virtual reality test, dichoptic stimulus presentation was used, whereby each eye received different visual information simultaneously to achieve binocular dissociation within the immersive environment. Participants used a handheld controller to align a virtual pointer with the sequentially presented stimuli. Each position was confirmed by maintaining alignment and pressing the trigger button. The system automatically recorded deviations in prism diopters and gaze coordinates.

Stimuli were presented in a predefined sequence covering the nine diagnostic gaze positions, ensuring consistency across participants and between repeated assessments.

2.6. Data Structure and Variables

The primary outcome variable was the point-by-point deviation measured in prism diopters at each of the nine diagnostic gaze positions for both assessment methods.

Each gaze position was analyzed as an individual observation. Therefore, each participant contributed multiple data points across gaze positions, methods (conventional Hess-Lancaster vs. Dicopt Pro), repetitions, and eyes, reflecting the repeated-measures structure of the dataset.

As a result, the dataset had a hierarchical structure, with measurements nested within gaze positions, repetitions, and methods, and ultimately within participants. Although gaze positions were treated as separate observations, the non-independence of measurements obtained from the same participant was accounted for in the statistical analysis.

This approach allowed evaluation of:

- agreement between methods at the level of individual gaze positions,
- within-method repeatability, and
- the influence of gaze position on measurement variability.

No aggregation of measurements was performed prior to analysis, preserving the full data structure. This approach ensured that all available measurements were retained in the analysis without loss of information.

2.7. Statistical Analysis

Statistical analysis was performed using SPSS (IBM Corp., Armonk, NY, USA) and R (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $p < 0.05$.

Continuous variables were summarized as mean \pm standard deviation (SD). The normality of the differences between methods was assessed using the Shapiro–Wilk test.

To account for the hierarchical structure of the dataset and repeated measurements within participants, method and gaze position were included as fixed effects, and participant was included as a random effect.

2.7.1. Agreement Analysis

Agreement between the conventional Hess-Lancaster test and the VR-based Hess-Lancaster assessment performed with Dicopt Pro was primarily evaluated using Bland–Altman analysis and concordance correlation coefficients (CCC). CCC values were interpreted as indicators of overall concordance between methods, combining measures of correlation and agreement.

Mean differences (bias) and 95% limits of agreement (LoA; mean difference \pm 1.96 SD) were calculated.

As multiple observations per participant were included, Bland–Altman plots were constructed using all paired measurements across gaze positions. Agreement analyses were conducted at the measurement level using paired point-coordinate observations and were interpreted considering the repeated-measures structure of the dataset.

Regression analysis of the differences against the mean values was performed to assess proportional bias.

2.7.2. Repeatability Analysis

Within-method repeatability was assessed separately for each method using intraclass correlation coefficients (ICC). A two-way mixed-effects model with absolute agreement for single measurements was applied. The 95% confidence intervals for ICC values were also calculated. ICCs were calculated separately for each method and eye using paired repeated point-coordinate measurements obtained during the same session. Repeatability was also assessed by comparing the two repeated measurements within each method at each gaze position.

2.7.3. Comparison Between Methods

Paired comparisons between methods were performed using paired t-tests for normally distributed data or Wilcoxon signed-rank tests otherwise. These analyses were considered complementary to agreement analysis, as statistical significance does not necessarily reflect clinical agreement.

2.7.4. Mixed-Effects Modeling

To account for the hierarchical structure of the data and repeated measurements within participants, linear mixed-effects models were applied using deviation values as the dependent variable. Method, eye, gaze position/component, and the method-by-eye interaction were included as fixed effects, while participant was included as a random effect. This model was used as a complementary analysis to evaluate systematic method-related differences while accounting for within-subject clustering.

2.7.5. Handling of Measurement Scale

As both systems were standardized to the same measurement scale (5 prism diopters per grid unit), no data transformation or conversion factors were required prior to analysis.

3. Results

3.1. Sample Characteristics

A total of 52 participants were included in the final analysis. The mean age was 58.0 ± 20.0 years. The sample included 33 female and 19 male participants.

3.2. Agreement Between Conventional Hess-Lancaster and VR-Based Hess-Lancaster Assessment

Bland–Altman analysis was performed to evaluate agreement between the conventional Hess-Lancaster test and the VR-based Hess-Lancaster assessment performed with Dicopt Pro using deviation measurements (Figure 3). For each eye, the analysis included 936 paired point-coordinate measurements, corresponding to 52 participants assessed across nine diagnostic gaze positions and two coordinate components. Global agreement results are summarized in Table 1.

For the right eye, the mean bias was 1.22 prism diopters, with 95% limits of agreement ranging from -17.46 to 19.89 prism diopters. For the left eye, the mean bias was 0.10 prism diopters, with 95% limits of agreement ranging from -18.44 to 18.64 prism diopters. The mean CCC was 0.57 for both eyes, and mean absolute differences were 4.78 and 4.59 prism diopters for the right and left eyes,

respectively. Overall, these findings indicate limited systematic bias, although some variability was observed at the individual measurement level. The complete point-by-point agreement analysis by eye, diagnostic gaze position, and coordinate component is provided in Supplementary Table S1.

As a sensitivity analysis, agreement was also evaluated after aggregating point-coordinate measurements at the participant-eye level (Supplementary Table S3). This analysis showed similar mean differences between methods, with narrower limits of agreement and higher CCC values, supporting the robustness of the main findings while reducing the influence of within-subject clustering.

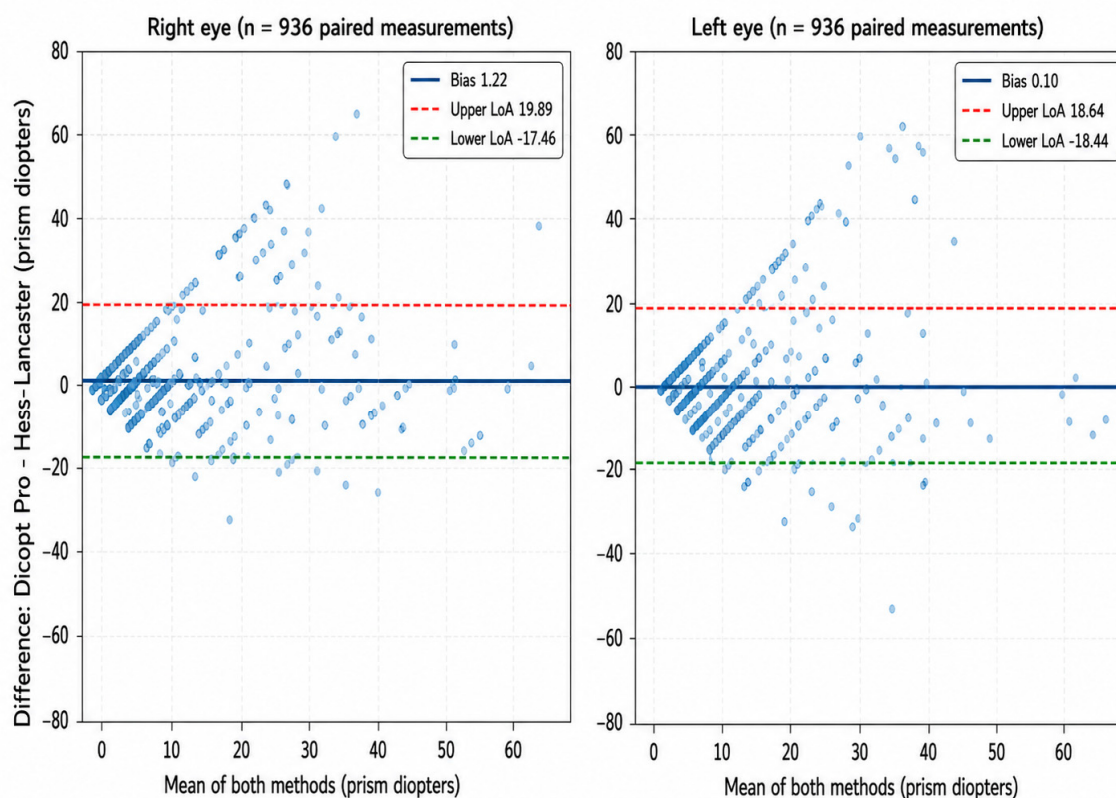


Figure 3. Bland–Altman plots showing point-by-point agreement between the conventional Hess-Lancaster test and VR-based Hess-Lancaster assessment performed with Dicopt Pro for the right and left eyes. Solid lines indicate the mean difference between methods, and dashed lines indicate the 95% limits of agreement.

Table 1. Global agreement between the conventional Hess-Lancaster test and the VR-based Hess-Lancaster assessment performed with Dicopt Pro using individual measurements.

Eye	Mean CCC	Mean difference (bias), Δ	95% LoA, Δ	Median difference, Δ	Mean absolute difference, Δ
Right eye	0.57	1.22	-17.46 to 19.89	0.00	4.78
Left eye	0.57	0.10	-18.44 to 18.64	-0.06	4.59

3.3. Agreement According to Diagnostic Gaze Position

Figure 4 shows the point-by-point agreement between the conventional Hess-Lancaster test and Dicopt Pro according to diagnostic gaze position and coordinate component. Overall, mean

differences between methods were generally close to zero in most gaze positions, suggesting limited systematic bias across the visual field. However, the width of the limits of agreement varied depending on the gaze position, indicating that agreement was not completely uniform across all diagnostic positions.

Greater variability was observed in some oblique and horizontal gaze components, whereas several vertical and central components showed narrower limits of agreement. This pattern suggests that the agreement between both methods may depend partly on the gaze direction assessed, with some positions showing more stable correspondence than others.

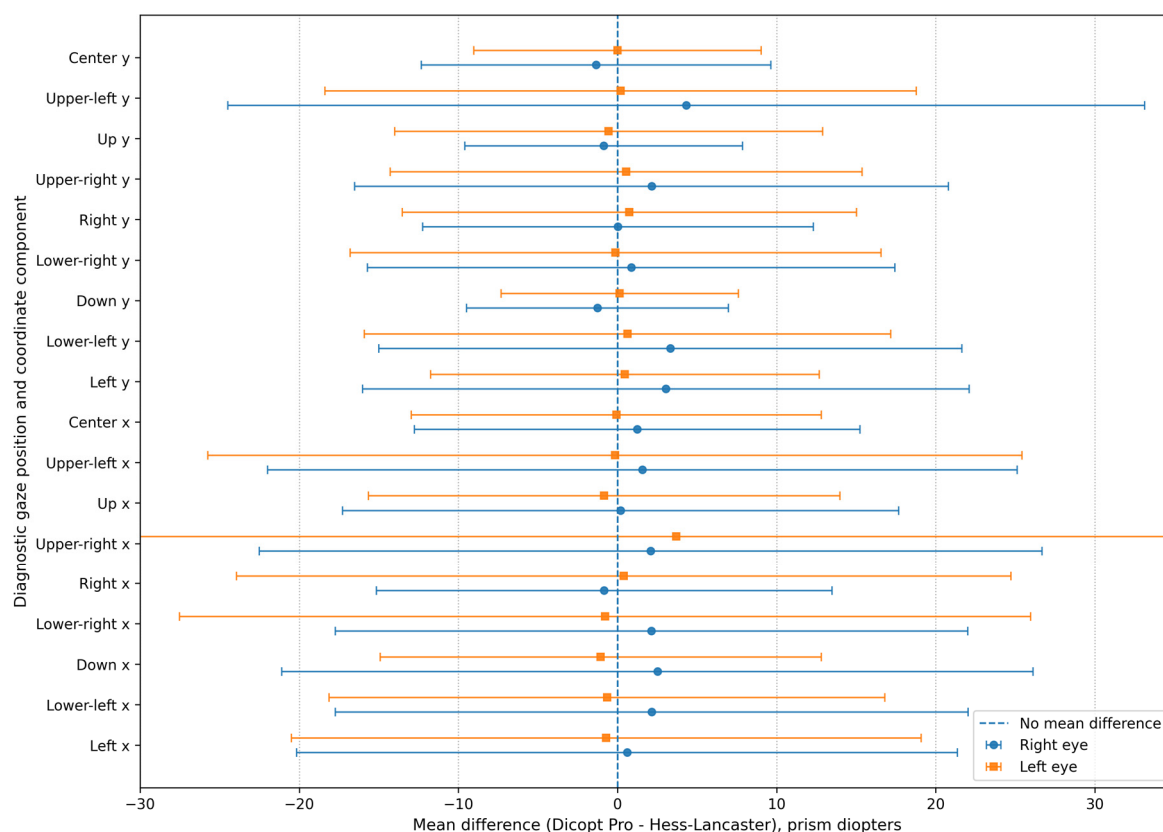


Figure 4. Forest plot of point-by-point agreement between the conventional Hess-Lancaster test and VR-based Hess-Lancaster assessment performed with Dicopt Pro.

Agreement according to diagnostic gaze position is summarized in Table 2 and illustrated in Figure 4. Table 2 presents the positions showing the highest and lowest agreement, while the full point-by-point results are reported in Supplementary Table S1. The highest agreement was observed for the lower gaze y-axis component, followed by upper and central gaze components. In contrast, lower CCC values were observed in some oblique and horizontal components, particularly in the upper-left and lower-left gaze positions.

Table 2. Selected diagnostic gaze positions showing higher and lower agreement between methods.

Gaze position/component	Mean CCC	Mean difference, Δ	Median difference, Δ	Mean absolute difference, Δ
Lower gaze, y-axis	0.84	-0.56	0.00	2.06
Upper gaze, x-axis	0.73	-0.32	0.00	4.36
Upper gaze, y-axis	0.72	-0.72	0.00	2.90

Central gaze, y-axis	0.70	-0.67	0.00	3.09
Right gaze, y-axis	0.70	0.38	0.00	3.06
Upper-left gaze, x-axis	0.44	0.70	-0.50	7.06
Left gaze, x-axis	0.43	-0.05	0.00	5.64
Upper-left gaze, y-axis	0.41	2.26	0.00	6.40
Lower-left gaze, y-axis	0.36	1.98	0.00	4.18

3.4. Within-Method Repeatability

Within-method repeatability was assessed in the subset of participants with repeated measurements available. Repeated conventional Hess-Lancaster measurements were available for 16 participants, whereas repeated Dicopt Pro measurements were available for 14 participants. Repeatability analysis is shown in Figure 5 and summarized in Table 3. ICC values were calculated using paired repeated point-coordinate measurements, whereas mean session-to-session differences and mean absolute differences were summarized at the participant-eye level. The complete repeatability analysis by method, eye, diagnostic gaze position, and coordinate component is provided in Supplementary Table S2.

The conventional Hess-Lancaster test showed small directional differences between the first and second measurements in both eyes. Dicopt Pro also showed relatively small between-session differences overall, although repeatability was less uniform, with lower repeatability for the right eye. Overall, ICC values indicated moderate-to-good within-method repeatability, although confidence intervals should be considered when interpreting the precision of these estimates.

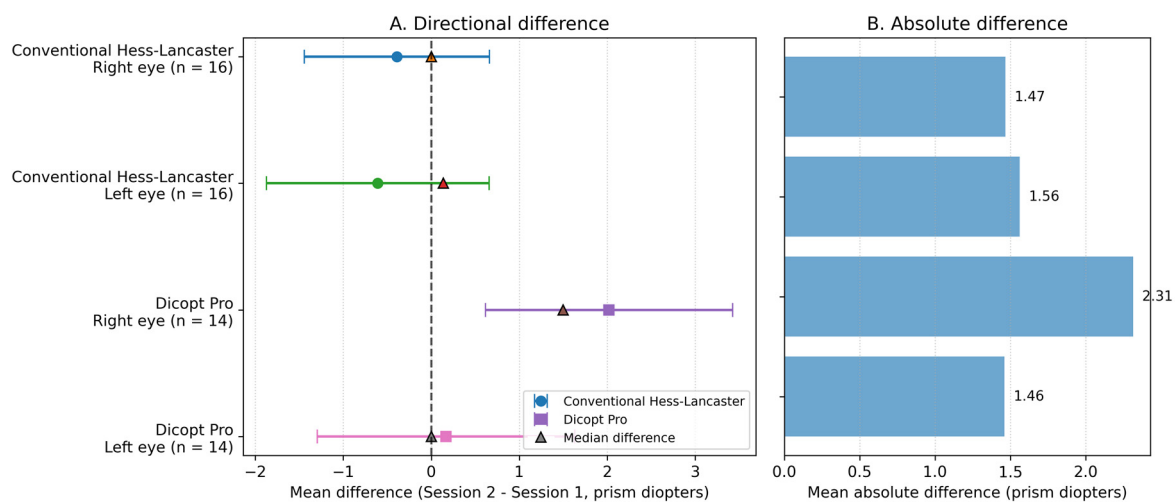


Figure 5. Within-method repeatability analysis for paired measurements.

Table 3. Within-method repeatability based on participant-level mean point-coordinate deviation measurements.

Method	Eye	<i>n</i>	Mean S1, Δ	Mean S2, Δ	Mean difference S2-S1, Δ	Mean absolute difference, Δ	p-value	ICC(A,1)	95% CI ICC
Conventional Hess-Lancaster	Right eye	16	4.60	4.21	-0.39	1.47	0.6948	0.73	0.61-0.82

Conventional Hess-Lancaster	Left eye	16	4.58	3.98	-0.61	1.56	1.0000	0.71	0.59–0.81
VR-based Hess-Lancaster assessment (Dicopt Pro)	Right eye	14	4.90	6.92	2.02	2.31	0.0159	0.62	0.51–0.74
VR-based Hess-Lancaster assessment (Dicopt Pro)	Left eye	14	6.15	6.31	0.17	1.46	0.5751	0.83	0.74–0.91

Note: ICC values correspond to a two-way mixed-effects model with absolute agreement for single measurements. The 95% confidence intervals were estimated using bootstrap resampling.

3.5. Mixed-Effects Model Analysis

Linear mixed-effects modeling was performed as a complementary analysis to evaluate systematic differences between methods while accounting for the repeated-measures structure of the dataset. Method, eye, gaze position/component, and the method-by-eye interaction were included as fixed effects, with participant as a random effect. The results of the model are summarized in Table 4.

The model showed a small but statistically significant method effect for the right eye, with Dicopt Pro measurements being on average 1.22 prism diopters higher than conventional Hess-Lancaster measurements ($\beta = 1.22 \Delta$; 95% CI: 0.52 to 1.91; $p = 0.001$). The method-by-eye interaction was also significant ($\beta = -1.11 \Delta$; 95% CI: -2.10 to -0.13; $p = 0.026$), indicating that the method-related difference was smaller for the left eye. Accordingly, the estimated method difference for the left eye was close to zero and not statistically significant ($\beta = 0.10 \Delta$; 95% CI: -0.59 to 0.80; $p = 0.772$). Overall, these findings support the Bland–Altman results, showing limited average systematic bias but eye-dependent differences between methods.

Table 4. Linear mixed-effects model evaluating systematic method-related differences.

Effect	Estimate, Δ	95% CI	p-value	Interpretation
Method effect, right eye	1.22	0.52 to 1.91	0.001	Dicopt Pro showed slightly higher values than conventional Hess-Lancaster
Method \times left eye interaction	-1.11	-2.10 to -0.13	0.026	The method-related difference was smaller for the left eye
Estimated method effect, left eye	0.10	-0.59 to 0.80	0.772	No significant systematic difference between methods

4. Discussion

The present study evaluated the agreement and within-method repeatability of deviation measurements obtained with the conventional Hess-Lancaster test and the VR-based Hess-Lancaster assessment performed with Dicopt Pro. The main finding was that the VR-based assessment showed moderate agreement with the conventional test across the nine-position framework. Mean differences between methods were small, particularly for the left eye, suggesting limited systematic bias. However, the limits of agreement showed some variability at the individual measurement level, indicating that point-by-point differences should be interpreted in relation to the clinical findings and the gaze position assessed.

The global concordance correlation coefficient was 0.57 for both eyes, indicating moderate agreement between the two methods. This result should be interpreted in the context of the characteristics of the conventional Hess-Lancaster test itself. Although the Hess-Lancaster screen remains a clinically valuable tool for mapping incomitant ocular deviations, it is not a fully objective reference standard. It depends on patient responses, examiner interaction, manual plotting, binocular dissociation, fixation stability, and test geometry. Therefore, discrepancies between the conventional and virtual methods may reflect not only limitations of the VR-based system but also the intrinsic variability of the traditional procedure.

The complementary mixed-effects model supported these findings by showing limited average systematic bias after accounting for within-subject clustering, although the method-related difference was slightly greater for the right eye than for the left eye.

Automated, video-based, eye-tracking, augmented reality, and virtual reality approaches have been proposed to reduce examiner-dependent variability and improve standardization in ocular alignment assessment. Previous studies have shown that head-mounted video systems, eye-tracking-enhanced Hess-Lancaster paradigms, automated strabismus devices, and immersive VR implementations can provide quantitative measurements of ocular deviation, although they also emphasize the importance of measurement variability, calibration, and repeatability when translating these technologies into clinical practice [7,8,19–21]. The present study extends this line of work by directly comparing a VR-based Hess-Lancaster assessment with the conventional clinical test under equivalent nine-position testing conditions.

The small mean bias observed in the Bland–Altman analysis suggests that the VR-based assessment did not systematically overestimate or underestimate point-by-point deviation values to a clinically dominant extent. However, the relatively wide limits of agreement indicate that individual differences should be interpreted according to the clinical context, gaze position, and purpose of the examination, particularly when measurements are used for surgical planning, longitudinal monitoring, or detailed characterization of incomitant deviations. This interpretation is consistent with methodological guidance emphasizing that agreement and repeatability should be interpreted in terms of measurement error and clinical acceptability rather than statistical significance alone [18,22]. Previous validation work using VR-based visual assessment has similarly emphasized the importance of considering device-specific scaling, systematic offsets, and measurement conditions when interpreting quantitative outputs [23].

Because each grid unit corresponded to approximately 5 prism diopters in both systems, differences below one grid unit may be interpreted cautiously as being within the resolution of the test. However, the clinical relevance of a given difference depends on gaze position, deviation magnitude, diagnosis, follow-up purpose, and whether the measurement is used for surgical planning. Therefore, rather than applying a single universal MCID, agreement was interpreted in relation to the 5Δ grid-unit resolution and the clinical context of ocular motility assessment.

Agreement was not uniform across diagnostic gaze positions. Higher agreement was observed in some central and vertical components, whereas lower CCC values and larger absolute differences were found in some oblique and horizontal components, particularly in upper-left and lower-left gaze positions. This position-dependent variability is clinically relevant because incomitant deviations are inherently gaze-dependent. Differences between methods may become more evident in eccentric gaze positions, where factors such as head positioning, perceived target location, spatial calibration, headset alignment, response strategy, or fixation stability may have a greater impact.

Although both systems were standardized to the same prism-diopter scale, the conventional test was performed at a physical distance of 60 cm with head stabilization, whereas the virtual assessment was performed at a simulated distance of 90 cm within a head-mounted display environment. This difference may have modified proximal convergence demand and contributed to variability in eccentric or oblique gaze positions, where spatial localization, vergence demand, headset alignment, and perceived target position may be more sensitive to test geometry. Previous comparisons between

dissociative screen tests also support the influence of viewing distance and experimental configuration on recorded deviation values [5].

The present findings also align with previous VR-based studies of ocular misalignment. Nesaratnam et al. reported that a virtual reality-based test of ocular misalignment was feasible when compared with the traditional Lees screen but also emphasized the need for further validation before clinical implementation [24]. More recently, Wang et al. evaluated a virtual reality system for automated strabismus measurement and diagnosis using Bland–Altman plots and ICC analysis, showing the potential of VR-based methods but also reporting systematic differences between VR and manual measurements [25]. These studies support the idea that VR technology can provide a controlled and scalable environment for ocular alignment assessment, while also highlighting that device-specific calibration, testing distance, response modality, and patient-related factors remain important sources of variability.

Within-method repeatability was moderate to good for both approaches, with ICC values of 0.73 and 0.71 for the conventional Hess-Lancaster test and 0.62 and 0.83 for Dicopt Pro, interpreted according to methodological recommendations for reliability studies [26]. However, Dicopt Pro showed lower repeatability for the right eye, consistent with the larger session-to-session difference observed in that condition. This eye-dependent variability may reflect learning or adaptation to the virtual environment, controller response variability, fixation instability, clinical heterogeneity, headset alignment, spatial calibration, or the limited size of the repeatability subsample.

Because the order of gaze-position presentation within each test followed a predefined sequence, a sequential learning or adaptation effect cannot be completely excluded. Similarly, manual controller use may have introduced response variability or fatigue-related effects, particularly during repeated VR measurements. Therefore, this finding should be interpreted as preliminary and hypothesis-generating rather than as evidence of software-related instability and should be confirmed in larger repeated-measures samples.

An important strength of this study is that both methods were compared using the same classical diagnostic gaze positions, preserving the clinical structure of the Hess-Lancaster test. This is relevant because several digital or automated approaches focus mainly on primary position or simplified gaze conditions. By maintaining the nine-position framework, the present study provides a more clinically meaningful comparison for incomitant deviations and paralytic or restrictive motility disorders. In addition, the use of Bland–Altman analysis, CCC, limits of agreement, and ICC provides a comprehensive methodological approach to evaluate agreement and repeatability.

The study also has limitations. First, the conventional Hess-Lancaster test was used as the clinical reference, although it is itself subject to variability. Second, the repeatability analysis was performed in a smaller subset of participants, which limits the precision of the repeatability estimates. In addition, because multiple point-coordinate measurements were obtained from each participant, agreement estimates were interpreted at the measurement level and in the context of the repeated-measures structure of the dataset. Mixed-effects models were included as complementary analyses to account for within-subject clustering. One author had a commercial affiliation with Dicopt Pro; however, data interpretation and statistical analyses were performed collaboratively by the research team following predefined methodological criteria. Further independent studies may help confirm the reproducibility of these findings in other clinical settings.

Third, although the headset incorporates eye-tracking technology, eye tracking was not used for measurement acquisition in this protocol, and responses remained subject-dependent. Therefore, the present results reflect the performance of a VR-based Hess-Lancaster assessment rather than a fully automated eye-tracking measurement system. Fourth, although the sample included patients with suspected or diagnosed ocular motility disorders, the population could not be reliably stratified according to specific diagnosis, deviation type, or severity because some clinical subgroups were small and heterogeneous. Therefore, subgroup-specific agreement analyses were not performed. Future studies should evaluate whether agreement differs between comitant and incomitant strabismus, paralytic deviations, restrictive disorders, or small versus large deviations.

Future research should explore the integration of automated eye-tracking outputs into VR-based Hess-Lancaster assessment, the optimization of calibration procedures for eccentric gaze positions, and the clinical interpretation of vector-based deviation patterns. Recent evidence on automated and telemedicine-based strabismus assessment highlights the potential of digital tools to improve accessibility, reduce subjectivity, and support standardized documentation, while also emphasizing the need for robust validation against in-person clinical examination before widespread implementation [27,28]. In this context, the structured digital output generated by VR-based assessments may facilitate future computational analyses and data-driven interpretation of ocular motility patterns, but further refinement and validation remain necessary.

5. Conclusions

The virtual reality-based Hess-Lancaster assessment performed with Dicopt Pro showed small mean differences and moderate agreement with the conventional Hess-Lancaster test across diagnostic gaze positions. However, the relatively wide limits of agreement indicate that individual point-by-point measurements should be interpreted within the context of the complete clinical examination rather than as directly interchangeable values. Within-session repeatability was moderate to good for both approaches, although some variability was observed with the VR-based assessment. Overall, VR-based Hess-Lancaster assessment may provide a standardized digital approach for ocular motility evaluation, with potential advantages in data recording and test standardization, while further studies are needed to confirm its clinical performance in larger and diagnostically stratified samples.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Table S1: Full point-by-point agreement analysis between the conventional Hess-Lancaster test and Dicopt Pro according to diagnostic gaze position and coordinate component; Table S2: Full within-method repeatability analysis for point-by-point deviation measurements according to diagnostic gaze position and coordinate component; Table S3: Sensitivity analysis based on participant-eye aggregated measurements.

Author Contributions: Conceptualization, F.J.P.-M.; methodology, F.J.P.-M., A.P.-S. and D.L.L.; investigation, F.J.P.-M., A.P.-S., D.L.L. and R.G.-J.; formal analysis, F.J.P.-M. and R.B.-V.; data curation, F.J.P.-M., A.P.-S. and D.L.L.; writing—original draft preparation, F.J.P.-M.; writing—review and editing, F.J.P.-M., A.P.-S., D.L.L., R.G.-J., R.B.-V. and J.E.C.-S.; supervision, F.J.P.-M.; project administration, F.J.P.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of Hospital Clínico San Carlos, Madrid, Spain (approval code: 25/147-E; date of approval: March 6, 2025).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request. Additional point-by-point agreement and repeatability results are provided in the Supplementary Materials.

Acknowledgments: The authors would like to express their sincere gratitude to Dr. María Rosario Gómez de Liaño Sánchez, Head of the Ocular Motility Unit of the Department of Ophthalmology at Hospital Clínico San Carlos, Madrid, and Professor of Ophthalmology at the Complutense University of Madrid, for her invaluable support, expert guidance, and generous collaboration throughout the development of this study. Her clinical expertise in ocular motility and strabismus was of great value for the conception and completion of this work.

Conflicts of Interest: A.P.-S. is commercially affiliated with Dicopt Pro. The remaining authors declare no conflicts of interest. The role of A.P.-S. in the study was limited to data acquisition and technical support, and he was not involved in the final statistical analysis or interpretation of the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
BA	Bland–Altman
CCC	Concordance correlation coefficient
CI	Confidence interval
ICC	Intraclass correlation coefficient
LoA	Limits of agreement
ML	Machine learning
SD	Standard deviation
S1	First session
S2	Second session
VR	Virtual reality
Δ	Prism diopters

References

1. Von Noorden, G.K.; Campos, E.C. *Binocular Vision and Ocular Motility: Theory and Management of Strabismus*, 6th ed.; Mosby: St. Louis, MO, USA, **2002**.
2. Roper-Hall, G. The Hess screen test. *Am. Orthopt. J.* **2006**, *56*, 166–174. <https://doi.org/10.3368/aoj.56.1.166>
3. Roodhooft, J.M. Screen tests used to map out ocular deviations. *Bull. Soc. Belge Ophthalmol.* **2007**, *305*, 57–67.
4. Hess, W.R. I. Ein einfaches messendes Verfahren zur Motilitätsprüfung der Augen. *Z. Augenheilkd.* **2010**, *35*, 201–219. <https://doi.org/10.1159/000294050>
5. Dysli, M.; Fierz, F.C.; Rappoport, D.; Meier, T.S.; Landau, K.; Bockisch, C.J.; et al. Divergence bias in Hess compared to Harms screen strabismus testing. *Strabismus* **2021**, *29*, 1–9. <https://doi.org/10.1080/09273972.2020.1871382>
6. Bergamin, O.; Zee, D.S.; Roberts, D.C.; Landau, K.; Lasker, A.G.; Straumann, D. Three-dimensional Hess screen test with binocular dual search coils in a three-field magnetic system. *Invest. Ophthalmol. Vis. Sci.* **2001**, *42*, 660–667.
7. Weber, K.P.; Rappoport, D.; Dysli, M.; Schmückle Meier, T.; Marks, G.B.; Bockisch, C.J.; et al. Strabismus measurements with novel video goggles. *Ophthalmology* **2017**, *124*, 1849–1856. <https://doi.org/10.1016/j.ophtha.2017.06.020>
8. Orduna-Hospital, E.; Maurain-Orera, L.; Lopez-de-la-Fuente, C.; Sanchez-Cano, A. Hess Lancaster screen test with eye tracker: An objective method for the measurement of binocular gaze direction. *Life* **2023**, *13*, 668. <https://doi.org/10.3390/life13030668>
9. Aimola, L.; Lane, A.R.; Smith, D.T.; Kerkhoff, G.; Ford, G.A.; Schenk, T. Efficacy and feasibility of home-based training for individuals with homonymous visual field defects. *Neurorehabil. Neural Repair* **2014**, *28*, 207–218. <https://doi.org/10.1177/1545968313503219>
10. Martino Cinnera, A.; Verna, V.; Marucci, M.; Tavernese, A.; Magnotti, L.; Matano, A.; et al. Immersive virtual reality for treatment of unilateral spatial neglect via eye-tracking biofeedback: RCT protocol and usability testing. *Brain Sci.* **2024**, *14*, 283. <https://doi.org/10.3390/brainsci14030283>
11. Otero-Currás, C.; Povedano-Montero, F.J.; Bernárdez-Vilaboa, R.; Rojas, P.; González-Jiménez, R.; Martínez-Florentín, G.; et al. Virtual reality-based dichoptic therapy in acquired brain injury: Functional and symptom outcomes. *J. Clin. Med.* **2026**, *15*, 1004. <https://doi.org/10.3390/jcm15031004>
12. Ting, D.S.W.; Pasquale, L.R.; Peng, L.; Campbell, J.P.; Lee, A.Y.; Raman, R.; et al. Artificial intelligence and deep learning in ophthalmology. *Br. J. Ophthalmol.* **2019**, *103*, 167–175. <https://doi.org/10.1136/bjophthalmol-2018-313173>

13. McDonald, M.A.; Tayebi, M.; McGeown, J.P.; Kwon, E.E.; Holdsworth, S.J.; Danesh-Meyer, H.V. A window into eye movement dysfunction following mTBI: A scoping review of magnetic resonance imaging and eye tracking findings. *Brain Behav.* **2022**, *12*, e2714. <https://doi.org/10.1002/brb3.2714>
14. Duchowski, A. *Eye Tracking Methodology: Theory and Practice*; Springer: London, UK, **2003**.
15. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
16. Startsev, M.; Agtzidis, I.; Dorr, M. 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behav. Res. Methods* **2019**, *51*, 556–572. <https://doi.org/10.3758/s13428-018-1144-2>
17. Ibragimov, B.; Mello-Thoms, C. The use of machine learning in eye tracking studies in medical imaging: A review. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 3597–3612. <https://doi.org/10.1109/JBHI.2024.3371893>
18. Bartlett, J.W.; Frost, C. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. *Ultrasound Obstet. Gynecol.* **2008**, *31*, 466–475. <https://doi.org/10.1002/uog.5256>
19. Grudzińska, E.; Durajczyk, M.; Grudziński, M.; Marchewka, Ł.; Modrzejewska, M. Usefulness assessment of automated strabismus angle measurements using innovative Strabiscan device. *J. Clin. Med.* **2024**, *13*, 1067. <https://doi.org/10.3390/jcm13041067>
20. Mehringer, W.; Wirth, M.; Risch, F.; Roth, D.; Michelson, G.; Eskofier, B. Hess screen revised: How eye tracking and virtual reality change strabismus assessment. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2021**, *2021*, 2058–2062. <https://doi.org/10.1109/EMBC46164.2021.9631002>
21. Nixon, N.; Thomas, P.B.M.; Jones, P.R. Feasibility study of an automated strabismus screening test using augmented reality and eye-tracking (STARE). *Eye* **2023**, *37*, 3609–3614. <https://doi.org/10.1038/s41433-023-02566-0>
22. McAlinden, C.; Khadka, J.; Pesudovs, K. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic Physiol. Opt.* **2011**, *31*, 330–338. <https://doi.org/10.1111/j.1475-1313.2011.00851.x>
23. Cedrún-Sánchez, J.E.; Bernárdez-Vilaboa, R.; Sánchez-Alamillos, L.; Medina-Galdeano, M.; Otero-Currás, C.; Povedano-Montero, F.J. Relationship Between Humphrey Automated Perimetry and Virtual Reality-Based Perimetry: A Constant dB Offset and Normative Data. *Applied Sciences* **2026**, *16*, 1351. <https://doi.org/10.3390/app16031351>
24. Nesaratnam, N.; Thomas, P.; Vivian, A. Stepping into the virtual unknown: Feasibility study of a virtual reality-based test of ocular misalignment. *Eye* **2017**, *31*, 1503–1506. <https://doi.org/10.1038/eye.2017.97>
25. Wang, Y.M.; Liu, J.; Chen, W.W.; Jiang, M.X.; Fu, J. Accuracy of virtual reality in automated measurement and diagnosis of strabismus. *Digit. Health* **2024**, *10*, 20552076241308713. <https://doi.org/10.1177/20552076241308713>
26. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
27. Hartness, E.M.; Jiang, F.; Zamba, G.K.D.; Allen, C.; Bragg, T.L.; Nellis, J.; et al. Automated strabismus evaluation: A critical review and meta-analysis. *Front. Neurol.* **2025**, *16*, 1620568. <https://doi.org/10.3389/fneur.2025.1620568>
28. Wong, D.S.H.; Alsaif, A.; Bender, L. The role of telemedicine in strabismus assessment: A narrative review and meta-analysis. *Telemed. J. E Health* **2024**, *30*, e2240–e2255. <https://doi.org/10.1089/tmj.2024.0115>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.