

Assessment of the Repeatability and Reproducibility of Binocular Vision Measurements Using a Novel Virtual Reality Application: EMAA Pro

Brodin Anaïs^{1*}, Pierre Fantou², Thomas Didier³, Audrey Persillion³ and Enora Robic⁴

¹Clinical Research Officer, Company Eyesoft, France

²Clinical research assistant at Eyesoft, France

³CEO of Eyesoft, France

⁴Orthoptiste, Rennes, France

*Corresponding Author: Brodin Anaïs, Clinical Research Officer, Company Eyesoft, France, E-mail: anais.b@eyesoft.fr

Received Date: October 10, 2025 **Accepted Date:** October 29, 2025 **Published Date:** November 01, 2025

Citation: Brodin Anaïs, Pierre Fantou, Thomas Didier, Audrey Persillion, Enora Robic (2025) Assessment of the Repeatability and Reproducibility of Binocular Vision Measurements Using a Novel Virtual Reality Application: EMAA Pro. J Ophthalmol Open Access 9: 1-14

Abstract

Significance

Virtual reality is increasingly being used in the medical field, particularly in rehabilitation. Eyesoft has developed a virtual reality headset combined with an eye tracker (EMAA Pro) to treat binocular vision disorders.

Purpose

The main objective of this study is to assess the repeatability, reproducibility and safety of the measurements carried out with EMAA. The secondary objective is to report expected values for all tests.

Methods

The study involved 70 non-presbyopic adults aged 18-40 with normal binocular vision and visual acuity of at least 20/25. All participants completed two test sessions using EMAA with the same examiner, measuring stereo acuity, NPC, fusional vergences, and ocular deviation. A subset of 36 subjects repeated these tests with a different examiner to assess inter-examiner reliability.



© 2025. Brodin Anaïs, Pierre Fantou, Thomas Didier, Audrey Persillion, Enora Robic. This is an open access article published by JScholar Publishers and distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Results

Stereo acuity was evaluated using Kappa coefficients, showing poor repeatability/reproducibility values ($K=0.3$ ($p<0.001$) / 0.2 ($p=0.04$) and $K=0.4$ ($p<0.001$) / 0.2 ($p=0.04$)). For other measurements, intra-class coefficient (ICC) and Bland-Altman analyses revealed: strong repeatability and reproducibility for NPC (both at $ICC=0.8$ ($p<0.001$)), positive fusional vergences ($ICC=0.8-0.9$ ($p<0.001$) / 0.6 ($p<0.001$) - 0.5 ($p=0.002$)) and ocular deviation values for distance ($ICC=0.7-0.8/0.7$ ($p<0.001$)) and near ($ICC=0.8-0.7/0.6-0.8$ ($p<0.001$)). However, values for negative fusional vergences were poor ($ICC=0.4$ ($p<0.001$) - 0.2 ($p=0.18$) / 0.2 ($p=0.15$) - 0.1 ($p=0.22$)).

Conclusion

This study showed overall good repeatability and moderate reproducibility of measurements using EMAA.

Introduction

Traditionally, binocular vision has been assessed using conventional tools such as prisms, prism bars, fixation targets, and standardized tests. However, these methods present several limitations. Measurement outcomes are strongly influenced by both the testing procedure and the examiner's technique, leading to variable and sometimes inconsistent results. For instance, in the evaluation of fusional vergences, the different available tests are not interchangeable and may produce heterogeneous outcomes [1, 2]. Furthermore, repeated use of these manual tools can expose clinicians to the risk of musculoskeletal disorders [3].

In recent years, virtual reality (VR) has emerged as an innovative technology in the medical field, particularly in rehabilitation, where it supports and enhances clinical practice. Its immersive properties offer promising alternatives to traditional vision testing. VR-based interventions have shown significant improvements in binocular vision parameters among young adults with convergence insufficiency and accommodative dysfunctions [4]. VR has also demonstrated reliability and validity in the measurement of strabismus [5, 6] and stereoscopic vision [7], producing results comparable to conventional assessments and exhibiting good test-retest consistency. The integration of eye-tracking technology within VR systems has further advanced the field by allowing precise quantification of ocular movements, thereby increasing the objectivity of measurements [8]. Additionally, the inclusion of gamification elements enhances patient engagement and compliance during visual tasks [9].

Within this context, the company Eyesoft has developed EMAA Pro, a system combining VR and integrated eye-tracking technology for the assessment and rehabilitation of binocular vision disorders, both in clinical and tele-care. The main advantage of this device is its ability to conduct a comprehensive orthoptic evaluation using immersive and accurate digital measurements. However, to date, no data are available regarding the repeatability and reproducibility of this tool.

Therefore, the present study aims to evaluate these parameters for several binocular vision measures specifically, stereoscopic acuity, near point of convergence (NPC), ocular deviation at near and distance, and positive and negative fusional vergences (PFV and NFV). Furthermore, this study seeks to establish reference values for these parameters when measured using the EMAA Pro headset.

Materials and Methods

Study Population

This prospective study was conducted at Rennes University Hospital, following approval from the institution's Biomedical Research Ethics Committee (ethical clearance reference number n°24.141).

Participants were recruited according to the following inclusion criteria:

- Age between of 18-40 years
- Normal retinal correspondence

- Minimum binocular visual acuity (VA) of at least 20/25 on the Monoyer linear scale.

All enrolled participants met these criteria, and no exclusions were required during the course of the study period.

Study Process

Prior to enrollment, all potential participants take part in a baseline screening to verify eligibility criteria. The initial evaluation included assessment of stereoscopic vision using the Lang I test at 40 cm and measurement of binocular visual acuity (VA) with the Monoyer linear scale at 5 meters. Subjects demonstrating a binocular VA of 20/25 or better and a positive stereopsis result (i.e., presence of binocular vision) were eligible for inclusion. All participants meeting these criteria were enrolled in the study.

The experimental protocol consisted of multiple measurement sessions. In the first phase, all participants underwent two consecutive examinations (Test 1 and Test 2) performed by Operator 1. The following parameters were recorded:

- Age (years)
- Gender
- Refractive status (emmetropia, myopia, hyperopia)
- NPC in centimeters (cm)
- Stereoacuity in seconds of arc (")
- PFV and NFV in prism diopters (PD)
- Ocular deviation at distance and near in prismatic diopters (PD).

Refractive errors were categorized into three groups: emmetropia, myopia and hyperopia. For subjects with astigmatism, spherical equivalent calculations were calculated to determine the appropriate categorization. In the second phase, the same measurements were repeated by Operator 2 during Tests 3 and 4, conducted a few days later, for a subset of the study population.

Materials

This study utilized the EMAA Pro application's Eval module, operated on a Pico Neo 3 Eye VR headset equipped with integrated Tobii eye-tracking technology (sampling rate: 30 Hz). To ensure methodological consistency, identical software versions and hardware configurations were maintained across all testing sessions.

A standardized testing protocol was implemented. Each participant completed the full assessment sequence twice to evaluate repeatability, in the following order: stereoscopic vision, NPC, PFV, NFV, and ocular deviation at both distance and near.

The system operates on an internet-based platform, requiring stable network connectivity. The integrated eye-tracking module continuously records ocular movements throughout each test. Data acquisition is fully automated, with results stored within pre-established individual participant profiles, enabling immediate post-assessment access for analysis. Participants were seated in a standard examination room and fitted with the VR headset, which provided complete visual isolation from the external environment. Prior to testing, the headset underwent calibration, during which participants fixated on a point alternately displayed at the four corners of the visual field inside the headset. All test procedures were explained verbally and supplemented by standardized written instructions displayed within the headset interface. Testing began with the stereoscopic vision assessment, performed using a random-dot stereotest that required participants to identify geometric shapes (square, triangle, or circle) across disparity levels ranging from 1000 to 30 arcseconds. The finest level successfully completed was recorded as the stereoscopic threshold. Subsequently, NPC was measured by asking participants to maintain fusion on a central target until the onset of diplopia. The NPC value corresponded to the closest point at which binocular convergence was maintained, up to a minimum limit of 5 cm. PFV and NFV were then assessed at a viewing distance of 2 meters. Participants were instructed to maintain single binocular vision while focusing on a white target. NFV (up to 20 PD) was evaluated first, followed by PFV (up to 80 PD). The fusion break point for each was automatically determined by the eye-tracking system. Finally,

ocular deviation was measured at near (40 cm) and distance (4 m). Participants fixated on a star-shaped target while an alternate prism cover test was performed to quantify deviation in prism diopters (PD).

Statistical Analysis

Test results were recorded using an Excel file with subsequent data pseudonymization. Demographic comparisons employed Student's t-test for age analysis and Chi-squared tests for gender and refractive error distributions. Spearman correlation analyses were assessed for quantitative measurements.

Repeatability was assessed by comparing Test 1 vs. Test 2 and Test 3 vs. Test 4 (intra-examiner), while reproducibility was assessed by comparing Test 1 vs. Test 3 and Test 2 vs. Test 4 (inter-examiner). Intraclass correlation coefficients (ICC) were used to determine the agreement of quantitative measures (NPC, PFV, NFV, and ocular deviation) following the methodology reported in [10].

Agreement between measurements was evaluated using Bland-Altman plots. The analysis included limits of agreement (LoA), defining the acceptable range of differences between repeated measurements to assess measurement variability and systematic error between testing methods. For categorical measurements (stereoscopic acuity), Cohen's Kappa coefficient assessed intra- and inter-observer agreement, providing reliability and reproducibility measures for subjective classifications according to [11].

Expected values of EMAA's measurements were presented as mean (with standard deviation), 95% CI, median, quartiles, and minimum and maximum values.

To evaluate the safe use of EMAA, investigators documented all adverse events occurring during VR exposure. These events were quantified as percentages of total.

Results

Population Demographics

This study involved 70 participants aged 18-40 (mean 25 ± 7 years), with 55 women and 15 men (sex ratio 3:6). All participants (Population 1) completed Tests 1 and 2. A subset of 36 participants (Population 2) completed Tests 3 and 4 with a different operator. Population 2 had a mean age of 23 ± 5 years, with 31 women and 5 men (sex ratio 6:2). Statistical analysis showed no significant demographic differences between the populations (Table 1).

All data were obtained for each participant based on their population.

Correlation coefficients were calculated for all quantitative measurements (Table 2). Strong and very strong correlations were observed for NPC and both near and far ocular deviation. For PFV, tests conducted for repeatability demonstrated very strong correlations, while those for reproducibility showed moderate correlations. However, all NFV tests showed lower correlations in both repeatability and reproducibility analyses.

Table 1: Characteristic Comparison between the Two Populations.

		Population 1 (n=70)	Population 2 (n=36)	P-value
Age (years)		25	23	0.11
Sex	Male	21	14	0.35
	Female	49	22	
Ametropia	Emmetropic	30	11	0.09
	Myopic	33	22	
	Hyperopic	7	3	

Table 2: Spearman Correlation Analysis of EMAA Pro Application Tests and Their Repeatability and Reproducibility

	Test 1-2		Test 3-4		Test 1-3		Test 2-4	
	r	p	r	p	r	p	r	p
NPC	0.8	<0.001	0.7	<0.001	0.7	<0.001	0.7	<0.001
PFV	0.8	<0.001	0.9	<0.001	0.5	0.002	0.4	0.009
NFV	0.4	<0.001	0.1	0.533	0.2	0.366	0.02	0.91
Ocular deviation at distance	0.7	<0.001	0.7	<0.001	0.6	<0.001	0.7	<0.001
Near ocular deviation	0.7	<0.001	0.8	<0.001	0.8	<0.001	0.6	<0.001

Repeatability

Repeatability was evaluated by comparing intra-examiner measurements, specifically between test sessions 1 and 2, and between sessions 3 and 4. Overall, a good level of agreement was observed, with negligible mean differences between repeated measurements for the NPC (0.03 ± 4.70 cm and 0.42 ± 2.89 cm), distance phoria (0.09 ± 1.71 PD and 0.33 ± 1.49 PD), and near phoria (0.13 ± 2.40 PD and 0.50 ± 3.90 PD). (Tab.3) Nevertheless, the limits of agreement (LoA), which reflect the range of measurement variability, should be carefully considered in clinical practice: ± 9.00 cm or ± 5.50 cm for NPC, ± 3.00 PD or ± 2.50 PD for

distance phoria, and ± 4.50 PD or ± 7.50 PD for near phoria. (Figure 1) Regarding PFV, intersession agreement was acceptable, although the mean differences between test sessions were larger (4.11 ± 15.40 PD and 1.64 ± 11.51 PD) (Table 3). Measurement variability was substantial, with a LoA of ± 30.00 PD or ± 22.00 PD, indicating that caution should be exercised when interpreting individual results in routine clinical settings. (Figure.1) For NFV, agreement was weak, despite small mean differences between sessions (0.69 ± 4.35 PD and 0.78 ± 4.93 PD) (Table 3). The LoA revealed considerable variability for this parameter (± 8.50 PD or ± 9.50 PD), suggesting that NFV measurements may be less reliable across repeated assessments (Figure 1).

Table 3: Intraclass Coefficient (Icc) of Repeatability of Emaa Pro Measurements

Measurements	Combination	ICC values	95% CI	P-values
NPC	Test 1 - Test 2	0.8	[0.69; 0.85]	<0.001
	Test 3 - Test 4	0.8	[0.61; 0.86]	<0.001
PFV	Test 1 - Test 2	0.8	[0.70; 0.86]	<0.001
	Test 3 - Test 4	0.9	[0.81; 0.93]	<0.001
NFV	Test 1 - Test 2	0.4	[0.25; 0.57]	<0.001
	Test 3 - Test 4	0.2	[-0.13; 0.41]	0.18
Far ocular deviation	Test 1 - Test 2	0.7	[0.60; 0.80]	<0.001
	Test 3 - Test 4	0.8	[0.68; 0.88]	<0.001
Near ocular deviation	Test 1 - Test 2	0.8	[0.74; 0.87]	<0.001
	Test 3 - Test 4	0.7	[0.46; 0.79]	<0.001

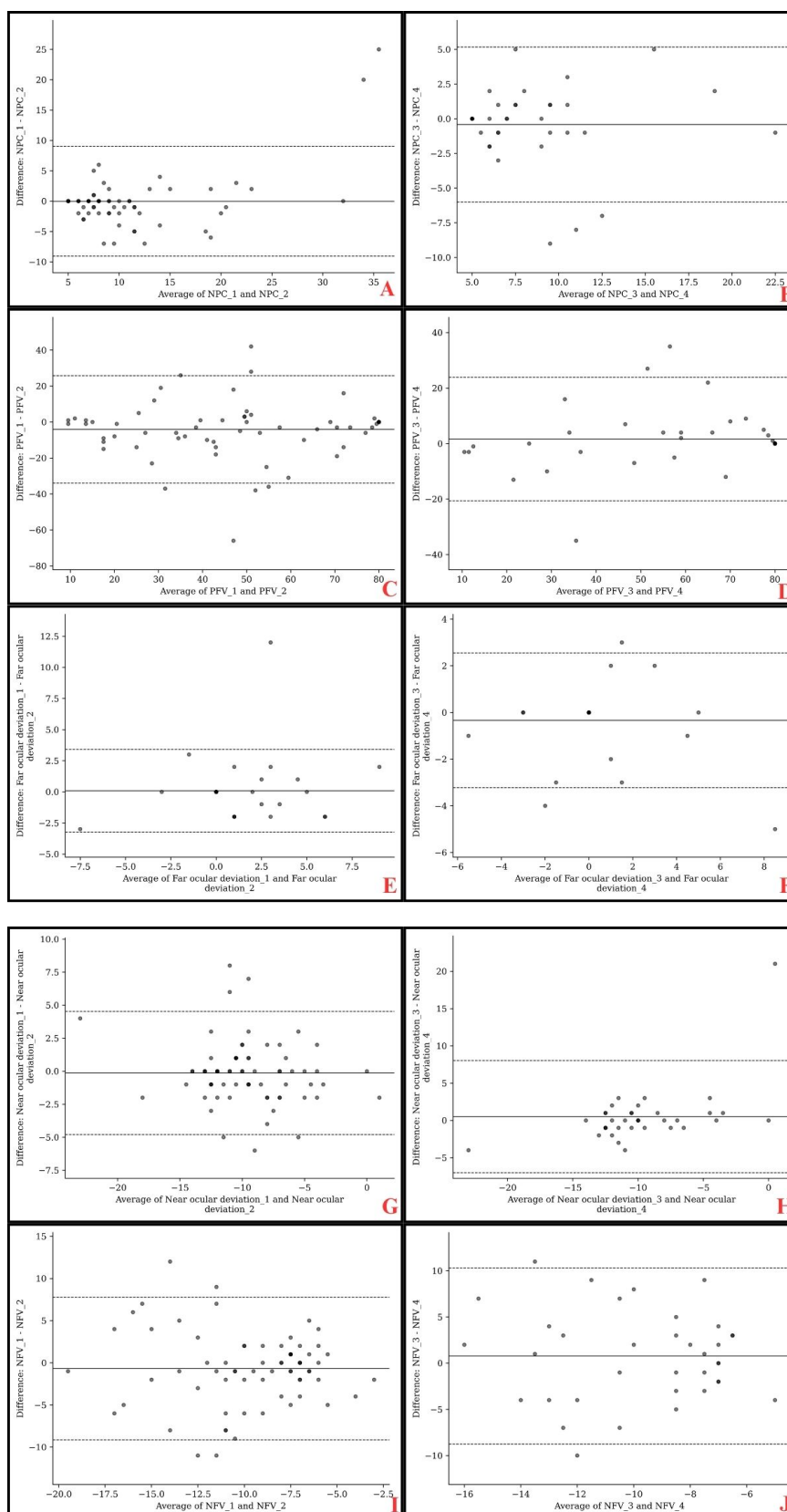


Figure 1: Bland-Altman plots comparing repeatability of EMMA measurements

A : NPC of Test 1-2 ; B : NPC of Test 3-4 ; C : PFV of Test 1-2 ; D : PFV of Test 3-4 ; E : Far ocular deviation of Test 1-2 ; F : Far ocular deviation of Test 3-4 ; G : Near ocular deviation of Test 1-2 ; H : Near ocular deviation of Test 3-4 ; I : NFV of Test 1-2 ; J : NFV of Test 3-4

Reproducibility

Inter-examiner reproducibility of EMAA Pro was evaluated by comparing measurements between Tests 1 and 3, and between Tests 2 and 4. Overall, good agreement was observed, with minimal mean differences for the NPC (0.28 ± 2.74 cm and 0.22 ± 3.00 cm) and for one of the near ocular deviation comparisons (0.22 ± 2.86 PD) (Table 4). However, measurement variability should be taken into account in clinical practice, as indicated by the LoA: ± 5.00 cm or ± 5.50 cm for NPC and ± 5.50 PD for the test 2–test 4 comparison of near ocular deviation (Figure 2). For distance ocular deviation (0.00 ± 1.60 PD and 0.14 ± 1.60 PD), PFV

(4.08 ± 21.75 PD and 1.72 ± 24.69 PD), and the Test 1–Test 3 comparison of near ocular deviation (0.06 ± 4.09 PD), agreement was moderate, with varying mean differences across parameters (Table 4). The corresponding LoA were also substantial and should be considered in clinical interpretation: ± 3.00 PD for distance ocular deviation, ± 47.5 PD for PFV, and ± 8.00 PD for the Test 1–Test 3 comparison of near ocular deviation (Figure 2). Finally, reproducibility was lower for NFV, which demonstrated poor agreement despite small mean differences (0.72 ± 5.00 PD and 1.00 ± 5.13 PD). (Table 4). Measurement variability was considerable for this parameter, with a LoA of ± 9.50 PD and ± 10.00 PD (Figure 2).

Table 4: Intraclass coefficient (ICC) of reproducibility of EMAA Pro measurements

Measurements	Combination	ICC values	95% CI	P-values
NPC	Test 1 - Test 3	0.8	[0.62; 0.86]	<0.001
	Test 2 - Test 4	0.8	[0.63; 0.86]	<0.001
PFV	Test 1 - Test 3	0.6	[0.42; 0.76]	<0.001
	Test 2 - Test 4	0.5	[0.23; 0.66]	0.002
NFV	Test 1 - Test 3	0.2	[-0.11; 0.43]	0.15
	Test 2 - Test 4	0.1	[-0.15; 0.39]	0.22
Far ocular deviation	Test 1 - Test 3	0.7	[0.59; 0.84]	<0.001
	Test 2 - Test 4	0.7	[0.57; 0.84]	<0.001
Near ocular deviation	Test 1 - Test 3	0.6	[0.43; 0.77]	<0.001
	Test 2 - Test 4	0.8	[0.60; 0.85]	<0.001

Regarding qualitative variables, stereoscopic vision was analyzed using the Kappa coefficient. The repeatability analysis revealed fair agreement between Test 1 and Test 2 ($K=0.3$; $p<0.001$), and slight agreement between Test 3 and Test 4 ($K=0.2$; $p=0.04$). For reproducibility, fair agreement was observed between Test 2 and Test 4 ($K=0.4$; $p<0.001$), while poor agreement was found between Test 1 and Test 3 ($K=0.2$; $p=0.04$), suggesting significant variability

between Observer 1 and Observer 2.

Expected Values

Expected values were calculated for each measurement session made with EMAA (Table 5). We focus primarily on the initial VR headset exposure measurements (Test 1), as these values are not influenced by learning or test-retest effects.

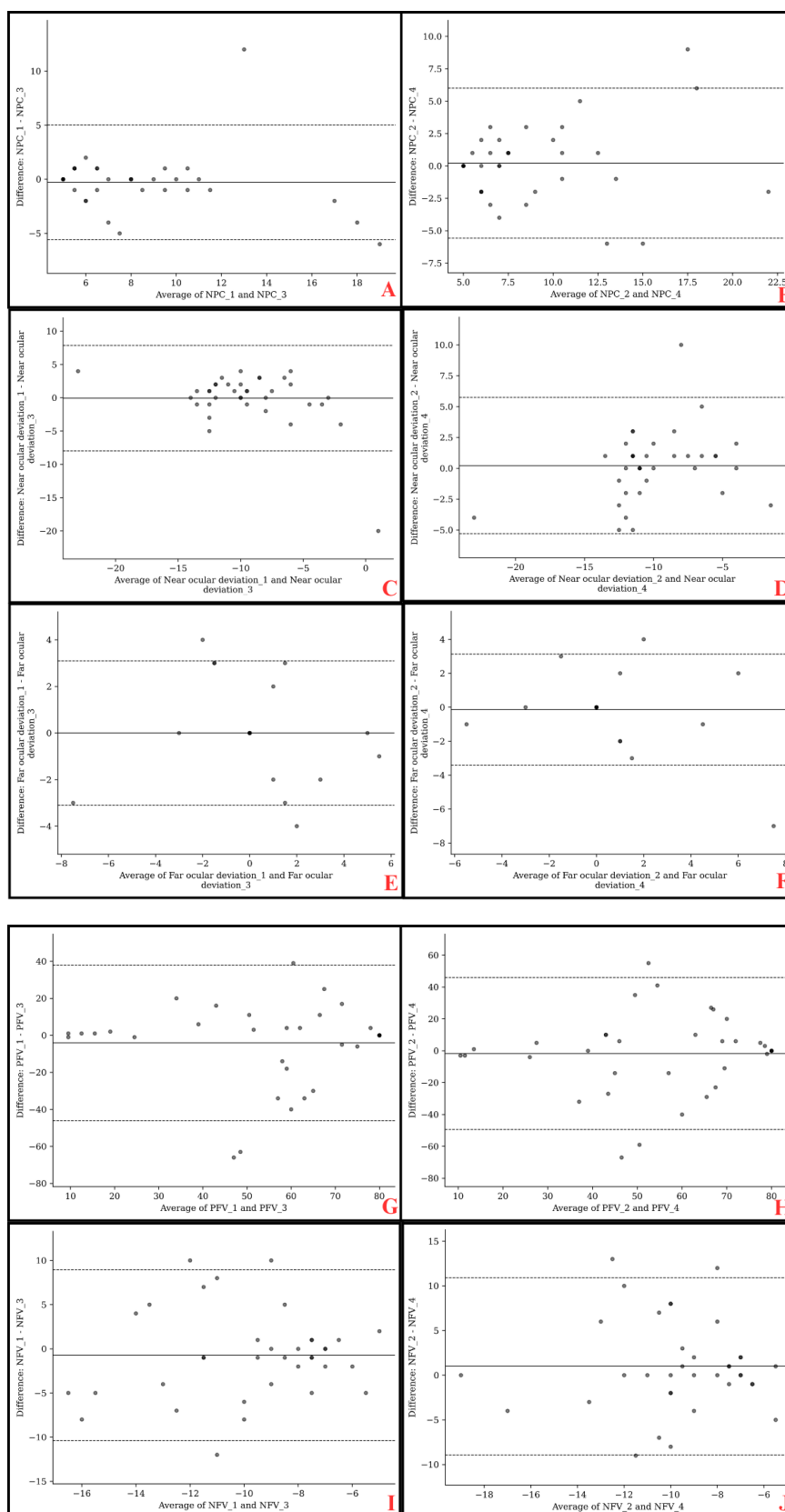


Figure 2: Bland-Altman Plots Comparing Repeatability of EMMA Measurements

A : NPC of Test 1-2 ; B : NPC of Test 3-4 ; C : Near ocular deviation of Test 1-2 ; D : Near ocular deviation of Test 3-4 ; E : Far ocular deviation of Test 1-2 ; F : Far ocular deviation of Test 3-4 ; G : PFV of Test 1-2 ; H : PFV of Test 3-4 ; I : NFV of Test 1-2 ; J : NFV of Test 3-4

Table 5: Excepted values of tests from EMAA Pro application

	Mean $\pm$ SD	95% CI	Min; Q1; median; Q3; max
Stereo acuity (in seconds of arc)	150.6 $\pm$ 143	[116.6;184.7]	(30; 45; 90; 215; 500)
NPC (in centimeters)	10.5 $\pm$ 6.6	[9;12.1]	(5; 6.5; 8.4; 11.5; 35.5)
PFV (in prism diopters)	51.2 $\pm$ 21.6	[46;56.3]	(10; 35.3; 50.8; 70.4; 80)
NFV (in prism diopters)	-9.7 $\pm$ 3	[-10.5;-9]	(-17.5; -11.4; -9; -7.5; -4)
Ocular deviation at distance (in prism diopters)	0.6 $\pm$ 2.1	[0.1;1.1]	(-6.5; 0; 0; 0.5; 9)
Near ocular deviation (in prism diopters)	-9.6 $\pm$ 3.7	[-10.4;-8.7]	(-23; -12; -10.3; -7.3; 1)

User Safety

Following each VR headset session, observers monitored patients for potential adverse events: headache, diplopia, vertigo, nausea or other adverse effects. Overall, 2.8% of participants (5 exposures) reported headaches as the sole symptom (3.6% in population 1 and 1.4% in population 2). Other symptoms were observed exclusively in population 2, with an ocular burning sensation reported in 1.4% of cases (1 exposure).

Discussion

The present study evaluated the repeatability and reproducibility of the EMAA Pro system for binocular vision measurements. For stereoacuity, repeatability agreement was fair ($\kappa = 0.3$ / $\kappa = 0.2$), reflecting relatively consistent assessments between observers. Reproducibility between Tests 1 and 3 showed similarly weak agreement ($\kappa = 0.2$). Performance varied substantially across stereoacuity levels. Previous studies assessing the repeatability of stereoacuity tests reported higher reliability, as reflected by the coefficient of repeatability (CoR). Among commonly used tests, the TNO test exhibited the lowest repeatability (CoR = $\pm 48''$), whereas the Frisby and Titmus tests demonstrated greater consistency (CoR = $\pm 10''$) [12, 13].

For the NPC, both repeatability and reproducibility analyses demonstrated good agreement (ICC = 0.7). In contrast, PFV yielded good repeatability (ICC = 0.7 / 0.8) but lower reproducibility (ICC = 0.6 / 0.5). NFV showed poor reliability for both repeatability (ICC = 0.4 / 0.2) and reproducibility (ICC = 0.2 / 0.1). These findings are comparable to those of Ma et al. [14], who reported PFV repeatability at 5 m and 33 cm (ICC = 0.81 / 0.81), but achieved

higher NFV values (ICC = 0.63 / 0.72) in participants with intermittent exotropia. This means that EMAA allows identical measurements to be obtained from the same person and with different examiners. But variability must also be taken into account because, for some measurements, it exceeds the prism bar step.

For ocular deviation at distance (4 m), moderate agreement was found for both repeatability (ICC = 0.7 / 0.8) and reproducibility (ICC = 0.7 / 0.7). Near ocular deviation (40 cm) showed moderate agreement for repeatability (ICC = 0.8 / 0.7) and reproducibility (ICC = 0.6 / 0.8). Recent studies assessing horizontal ocular deviation through subjective techniques have reported slightly higher reliability; however, methodological differences must be considered. EMAA measurements are conducted in an immersive virtual environment without accommodative demand, which may influence results. Facchin and Maffioletti (2021) observed comparable repeatability for distance (ICC > 0.8 at 3 m) and near (ICC > 0.9 at 40 cm) [15], while Anstice et al. reported similar repeatability using the von Graefe method (ICC = 0.8 at both 6 m and 40 cm), with reproducibility of ICC = 0.9 (distance) and ICC = 0.8 (near) [16].

Regarding stereoacuity, the mean result was 182.1" $\pm$ 191.7", with distribution by level as follows: 30" (30%), 60" (14.3%), 120" (30%), and 500" (25.7%). Piano et al. reported lower mean values for the TNO test (52" $\pm$ 25" or 77" $\pm$ 82") and Frisby test (55" $\pm$ 2" or 21" $\pm$ 3") [17]. The EMAA stereotest appears to underestimate stereoacuity compared with traditional methods, possibly due to the absence of monocular and binocular non-stereoscopic cues that can confound conventional tests, as demonstrated by

[18, 19].

For NPC, although no universal standard exists, the French Ophthalmological Society defines a normal range as 8–10 cm from the orbital rim [20–22]. Reported lower values in children (2.5 ± 2.5 cm and 3.0 ± 3.0 cm for accommodative targets; 4.5 ± 3.5 cm and 4.0 ± 4.0 cm for non-accommodative red/green targets). The mean NPC obtained in this study is higher (10.5 ± 8.1 cm). It should be noted that EMAA limits NPC measurement to 5 cm, whereas clinical testing may continue up to the base of the nose. Differences from previous studies may arise from the use of non-accommodative targets and the adult population tested in the present work.

The mean PFV was 48.4 ± 24.7 PD, considerably higher than values obtained with conventional instruments such as the Berens prism bar (18.1 ± 1.3 PD [23]; 23.3 ± 7.7 PD [24]) or synoptophore (30.9 ± 3.8 PD [23]). Mean NFV was -9.7 ± 4.0 PD, consistent with literature values using prism bar (-7.2 ± 0.4 PD [23]; -8.6 ± 1.9 PD [24]) and synoptophore (-8.4 ± 0.7 PD [23]). EMAA allows PFV measurement up to 80 PD, exceeding the 40–45 PD range typically achievable with prism bars. Importantly, EMAA quantifies smooth vergences, whereas prism bars measure step vergences, which may account for differences in expected values. Normative values for smooth vergences are PFV: 19 ± 8 PD (distance) and 21 ± 6 PD (near), NFV: -7 ± 3 PD (distance) and -21 ± 4 PD (near); for step vergences, PFV: 11 ± 7 PD (distance) and 19 ± 9 PD (near), NFV: -7 ± 3 PD (distance) and -13 ± 6 PD (near) [21]. Antona et al. [24] emphasized that these two methods should not be used interchangeably.

For ocular deviation, the mean at distance was 0.6 ± 2.4 PD, consistent with conventional tests (cover test: -0.2 ± 1.3 PD; von Graefe: 0.9 ± 1.9 PD; Maddox: 0.3 ± 2.6 PD) [25]. In contrast, near ocular deviation (-9.6 ± 3.8 PD) differed significantly from reference values, in contrast with Cantó-Cerdán's findings (cover test: -3 ± 4 PD; von Graefe: -3.5 ± 4.7 PD), with LoA of ± 2.97 PD (distance) and ± 6.74 PD (near) [26]. These discrepancies may arise because EMAA testing involves no accommodative vergence, which tends to reduce esophoria and increase exophoria. Nonetheless, LoA values were comparable at distance (± 3.3 PD) and

improved at near (± 4.6 PD).

For the third objective, a total of 212 VR sessions were recorded. Only five exposures (2.8%) resulted in transient headaches, and one participant (1.4%) reported mild eye burning. These symptoms are nonspecific and not exclusive to VR or binocular vision assessment. Prior studies have categorized visual discomfort into external symptoms (burning, irritation, tearing, dryness) caused by prolonged fixation or glare, and internal symptoms (pain, tension, headaches) linked to accommodative or binocular strain [27, 28]. Although cyberkinesis (nausea, disorientation, dizziness) has been documented during VR immersion [29, 30], such effects were not observed in this study. Some research suggests that VR exposure may induce excessive accommodation and reduced convergence, potentially leading to visual fatigue in young adults; however, the safety of conventional orthoptic testing remains poorly documented [31].

This study presents certain limitations, notably an age imbalance within the sample (50% of participants under 24 years) and the inclusion of a predominantly orthoptic student population, though most were unfamiliar with EMAA testing. Overall, the findings demonstrate that EMAA Pro provides binocular vision measurements with variable repeatability and reproducibility depending on the parameter assessed. The system's objective and automated design contributes to enhanced consistency through standardized conditions (stimulus characteristics, illumination, and eye-tracking-based data collection). Conducting repeated assessments using the same headset or different operators should yield more reproducible outcomes than traditional manual techniques.

This preliminary investigation also established reference values for EMAA Pro measurements in a limited cohort ($n = 70$, aged 18–40 years). The observed systematic differences compared with conventional tests warrant further validation studies on a larger and more diverse population to confirm these findings and refine normative data.

Conclusion

The results of this study demonstrate good repeatability and moderate reproducibility for binocular vision measurements obtained using the EMAA Pro system. Never-

theless, these findings should be interpreted with caution, considering the discrepancies observed between the measured values and those reported in the existing literature, as well as the multiple factors that can influence outcomes (e.g., testing distance, methodology, and measurement conditions).

Importantly, the EMAA platform enables the safe and objective assessment of several binocular vision parameters through integrated eye-tracking technology. While EMAA provides standardized and controlled measure-

ments, these should always be interpreted within the context of a comprehensive clinical evaluation and the patient's reported symptoms. Furthermore, the system offers the added advantage of facilitating rehabilitation of binocular vision disorders through its immersive virtual interface.

As this work represents a preliminary investigation, further research is warranted to establish normative data in a larger and more diverse population particularly including children and presbyopic adults to confirm the validity and generalizability of this measurement approach.

References

1. Feldman, Jerome M, Jeffrey Cooper, Pat Carniglia, et al. (1989) "Comparison of Fusional Ranges Measured by Risley Prisms, Vectograms, and Computer Orthopter." *Optometry and Vision Science : Official Publication of the American Academy of Optometry*. 66: 375–82.
2. Lança, Carla Costa, and Fiona J. Rowe (2019) "Measurement of Fusional Vergence: A Systematic Review." *Strabismus*. 27: 88–113.
3. Pontheaux, Olivia, Anne Gallasso (2020) "Objectivation Des Troubles Musculo Squelettiques (TMS) Chez Les Orthoptistes Français-Projet 2ECTO." *Revue Francophone d'Orthoptie*. 13: 139–47.
4. Li, Shijin, Angcang Tang, Bi Yang, Jianglan Wang, Longqian Liu (2022) "Virtual Reality-Based Vision Therapy versus OBVAT in the Treatment of Convergence Insufficiency, Accommodative Dysfunction: A Pilot Randomized Controlled Trial." *BMC Ophthalmology* 22.
5. Nixon, Nisha, Peter B.M. Thomas, and Pete R. Jones (2023) "Feasibility Study of an Automated Strabismus Screening Test Using Augmented Reality and Eye-Tracking (STARE)". 37: 3609–14.
6. Yeh, Po Han, Chun Hsiu Liu, Ming Hui Sun, et al. (2021) "To Measure the Amount of Ocular Deviation in Strabismus Patients with an Eye-Tracking Virtual Reality Headset." *BMC Ophthalmology*. 21: 1–8.
7. Denkinger, Sylvie, Maria Paraskevi Antoniou, et al. (2023) "The ERDS v6 Stereotest and the Vivid Vision Stereo Test: Two New Tests of Stereoscopic Vision." *Translational Vision Science & Technology*. 12: 1.
8. Brunyé, Tad T, Trafton Drew, Donald L Weaver, and Joann G Elmore (2019) "A Review of Eye Tracking for Understanding and Improving Diagnostic Interpretation." *Cognitive Research: Principles and Implications*. 4: 1–16.
9. Boon, Mei Ying, Lisa J. Asper, Peiting Chik, Piranaa Alagiah, Malcolm Ryan (2020) "Treatment and Compliance with Virtual Reality and Anaglyph-Based Training Programs for Convergence Insufficiency." *Clinical & Experimental Optometry*. 103: 870–76.
10. Koo, Terry K, and Mae Y Li (2016) "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine*. 15: 155–63.
11. Landis, J. Richard and Gary G. Koch (1977) "The Measurement of Observer Agreement for Categorical Data." *Biometrics*. 33: 159.
12. Antona, Beatriz, Ana Barrio, Isabel Sanchez, et al. (2015) "Intraexaminer Repeatability and Agreement in Stereoacuity Measurements Made in Young Adults." *International Journal of Ophthalmology*. 8: 374–81.
13. Mehta, Jignasa, and Anna O'Connor (2023) "Test Retest Variability in Stereoacuity Measurements." *Strabismus*. 31: 188–96.
14. Ma, Martin Ming Leung, Ying Kang, Mitchell Scheiman, et al. (2022) "Reliability of Step Vergence Method for Assessing Fusional Vergence in Intermittent Exotropia." *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*. 42: 913–20.
15. Facchin, Alessio, and Silvio Maffioletti (2021) "Comparison, within-Session Repeatability and Normative Data of Three Phoria Tests." *Journal of Optometry* 14: 263–74.
16. Anstice, Nicola S, Bianca Davidson, Bridget Field, et al. (2021) "The Repeatability and Reproducibility of Four Techniques for Measuring Horizontal Heterophoria: Implications for Clinical Practice." *Journal of Optometry*. 14: 275–81.
17. Piano, Marianne EF, Laurence P Tidbury, and Anna R O'Connor (2016) "Normative Values for Near and Distance Clinical Tests of Stereoacuity." *Strabismus*. 24: 169–72.
18. Cooper, Jeffrey, Joel Warshowsky (1977) "Lateral Displacement as a Response Cue in the Titmus Stereo Test." *American Journal of Optometry and Physiological Optics*. 54: 537–41.
19. Chopin, Adrien, Samantha Wenyan Chan, Bahia Guellai, et al. (2019) "Binocular Non-Stereoscopic Cues Can Deceive Clinical Tests of Stereopsis." *Scientific Reports*. 9:

- 1–10.
20. “Rapport SFO - Strabisme.” n.d (2025). 312.
21. Scheiman, Mitchell, and Bruce Wick (2020) “Clinical Management of Binocular Vision : Heterophoric, Accommodative, and Eye Movement Disorders.” Section I, 8.
22. Hussaindeen, Jameel Rizwana, Archayeeta Rakshit, Neeraj Kumar Singh, et al. (2017) “Binocular Vision Anomalies and Normative Data (BAND) in Tamil Nadu: Report 1.” *Clinical and Experimental Optometry* 100: 278–84.
23. Haque, Shania, Sonia Toor, David Buckley (2024) “Are Horizontal Fusional Vergences Comparable When Measured Using a Prism Bar and Synoptophore?” *The British and Irish Orthoptic Journal*. 20: 85.
24. Antona B, A Barrio, F Barra, E Gonzalez, I Sanchez (2008) “Repeatability and Agreement in the Measurement of Horizontal Fusional Vergences.” *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*. 28: 475–91.
25. Cebrian, Jose Luis, Beatriz Antona, Ana Barrio, Enrique Gonzalez, et al. (2014) “Repeatability of the Modified Thorington Card Used to Measure Far Heterophoria.” *Optometry and Vision Science : Official Publication of the American Academy of Optometry*. 91: 786–92.
26. Cantó-Cerdán, Mario, Pilar Cacho-Martínez, Ángel García-Muñoz (2018) “Measuring the Heterophoria: Agreement between Two Methods in Non-Presbyopic and Presbyopic Patients.” *Journal of Optometry*. 11: 153–59.
27. Pölönen, Monika, Toni Järvenpää, Beatrice Bilcu (2013) “Stereoscopic 3D Entertainment and Its Effect on Viewing Comfort: Comparison of Children and Adults.” *Applied Ergonomics*. 44: 151–60.
28. Sheedy, James E, John Hayes, Jon Engle (2003) “Is All Asthenopia the Same?” *Optometry and Vision Science : Official Publication of the American Academy of Optometry*. 80: 732–39.
29. Kennedy, Robert S, Julie Drexler, Robert C. Kennedy (2010) “Research in Visually Induced Motion Sickness.” *Applied Ergonomics*. 41: 494–503.
30. Kim, Hyun K, Jaehyun Park, Yeongcheol Choi, Mungyeong Choe (2018) “Virtual Reality Sickness Questionnaire (VRSQ): Motion Sickness Measurement Index in a Virtual Reality Environment.” *Applied Ergonomics*. 69: 66–73.
31. Mohamed Elias, Zulekha, Uma Mageswari Batumalai, Azam Nur Hazman Azmi (2019) “Virtual Reality Games on Accommodation and Convergence.” *Applied Ergonomics*. 81: 102879.

Submit your manuscript to a JScholar journal and benefit from:

- § Convenient online submission
- § Rigorous peer review
- § Immediate publication on acceptance
- § Open access: articles freely available online
- § High visibility within the field
- § Retaining the copyright to your article

Submit your manuscript at
<http://www.jscholaronline.org/submit-manuscript.php>