

Original Research

Assessing the Inter-Rater and Inter-Trial Reliability of the NeuroOptics Pupillary Light Response-3000 Pupillometer

Deborah A Jehu^{1a}, Lori A Bolgla², Samantha Armas¹, Forest Dutton²

¹ Interdisciplinary Health Sciences, Augusta University, ² Physical Therapy, Augusta University

Keywords: pupillometry, pupillary light reflex, pupillary light response, psychometrics

<https://doi.org/10.26603/001c.128047>

International Journal of Sports Physical Therapy

Vol. 20, Issue 2, 2025

Background

An automated pupillometer is a handheld device used to stimulate the pupillary light response (PLR) and track the entirety of the response from constriction to dilation. Pupillometers provide objective data that clinicians can use to identify and assess brain injury. The validity of these devices has been previously established; however, the inter-rater and inter-trial reliability are unknown.

Purpose

The purpose of this study was to assess the inter-rater and inter-trial reliability of the NeuroOptics PLR-3000 pupillometer device in measuring pupil size changes, constriction velocities, and dilation velocities. The authors hypothesized that inter-rater and inter-trial reliability would have intraclass correlation coefficients (ICC) greater than or equal to 0.70 for all PLR parameters.

Study

Design: Observational, reliability study

Methods: Forty-eight healthy adults (age 18-40 years) without a history of neurological injury, optical surgery, or cognitive impairment participated. Two independent raters used the NeuroOptics PLR-3000 to measure PLR parameters in the left and right eyes of each subject. Data for the average and individual trials of each PLR parameter were used to determine inter-rater and inter-trial reliability, respectively. Inter-rater and inter-trial reliability was evaluated using descriptive statistics, ICC, the standard error of measurement, Bland-Altman plots, and the minimal detectable change.

Results

Seven out of eight NeuroOptics 3000-PLR parameters demonstrated moderate-to-excellent inter-rater (ICC range 0.72-0.96) and good-to-excellent inter-trial reliability (ICC range 0.76-0.98). The 75% recovery time parameter exhibited moderate inter-rater (ICC range 0.64-0.67) and poor-to-moderate inter-trial (ICC range 0.41-0.65) reliability.

Conclusion

The NeuroOptics 3000-PLR demonstrated acceptable reliability in measuring initial and end pupil size, constriction and dilation velocity, and latency to change between different users and trials. However, the device exhibited unacceptable reliability when measuring

^a Corresponding Author:
Deborah A. Jehu, PhD
Assistant Professor
Department of Community & Behavioral Health Sciences
School of Public Health
The Graduate School
Augusta University
Augusta, GA 30912
djehu@augusta.edu

the time to 75% pupil size recovery. The device can be used in detecting and monitoring brain injury but should be limited to reliable measures only.

INTRODUCTION

A concussion can provoke changes in pupil size and the pupillary light response (PLR). These subtle, yet significant, changes have led healthcare providers to measure neurological injury with a pupillometer.¹⁻⁴ Pupillometers can provide insight regarding the location of neurological lesions and predict recovery trajectory following traumatic brain injury.⁵ The PLR is both a response and a visual reflex to the level of light sensed in the environment and serves as an accessible marker of the autonomic nervous system.^{6,7} The PLR provides a comprehensive manner to assess sympathetic and parasympathetic function.⁷ Specifically, the sympathetic pathway controls the eye muscles responsible for pupil dilation, while the parasympathetic pathway controls the eye muscles responsible for pupil constriction.⁸ The parasympathetic system causes eye constriction when a light stimulus is applied, and the sympathetic system causes eye dilation to the baseline state when the light stimulus is removed. Studying the static and dynamic properties of the PLR has emerged as an attractive field of interest given its ease of access, non-invasiveness, and insight into numerous neurological disorders and physiological states.^{6,9-11}

Historically, healthcare providers have used penlights to assess pupil symmetry and PLR. Concerns related to the use of penlights are low inter-rater reliability, higher error rates in prognosis, and the reduced ability to monitor the recovery of the PLR.^{2,5,6} Automated pupillometer systems have been developed and shown to be more accurate and reliable for examining the PLR.^{2,12} These devices are superior to manual observation because of their ability to monitor intracranial pressure, provide a prognosis following concussion, and assess cognitive load.^{1,6,11,13-16} At a minimum, automated devices provide readouts on the static (e.g., minimum and maximum pupil diameter) and dynamic (pupil constriction and dilation velocity) parameters.¹⁵ More advanced devices, such as the NeuroOptics Neurologic Pupil index (NPI)-200 and NPI-300, provide the NPI with a calculation that incorporates pupil size, constriction latency, constriction velocity, and dilation velocity. These devices are useful because they can compare scores to a normative database. More specifically, they provide a score range of 0-5 points, in which a score less than 3 points indicates abnormal pupil function.¹⁷ The NeuroOptics PLR-3000 (NO3000) better characterizes the PLR response as it provides the time from peak pupil constriction size to 75% of its baseline size, commonly known as the T-75 recovery time parameter (T75). Users of earlier models like NPI-300 could not obtain this parameter and had to extrapolate the graphical data to calculate T75.¹⁰

The T75 represents the sympathetic drive behind the dilation phase and is influenced by the amplitude of the light reflex.¹⁸ The larger the percent change from baseline to maximum dilation size results in more time needed for the pupil to constrict and return to baseline. Researchers have

reported longer T75 times in children with mild concussions and athletes with sport-related concussions compared to controls.^{14,19} While the T75 can discriminate between concussed and healthy groups, its reliability has not been extensively examined.^{14,19}

Establishing the psychometric properties of commonly used PLR systems is important to ensure they appropriately acquire meaningful information that aids in diagnosis, clinical prognosis, and research. In addition to concussions and traumatic brain injuries, pupillometers show promise in better understanding different neurological and chronic diseases such as Parkinson's and Alzheimer's.^{4,6,9} As these systems, particularly the hand-held automatic pupillometers, are increasingly integrated into common healthcare settings, the importance of verifying a device's robustness is critical.^{4,9,17,20} Importantly, devices have inherent differences due to their design, which may introduce measurement variability.⁴ NO3000 has established inter-trial reliability, but inter-rater reliability has not been established.²¹ Therefore, the purpose of our study was to establish the inter-rater reliability and confirm the inter-trial reliability of the NO3000 pupillometer among healthy adults. The authors hypothesized that inter-rater and inter-trial reliability for all PLR measures would have intraclass correlation coefficients (ICC) greater than or equal to 0.70, which has been deemed as acceptable reliability.²²

METHODS

PARTICIPANTS

Before subject recruitment, we conducted an *a priori* power analysis. Based on a minimum acceptable ICC of 0.70 and expected reliability ICC of 0.86, a two-tailed significance of $\alpha=0.05$, a power of 80%, and two raters, at least 39 subjects would be needed.²² Forty-eight healthy adults were recruited from Augusta University via word of mouth and email advertisement (25 males, age = 25.0 ± 4.7 y; 23 females, age = 25.3 ± 6.4 y). Eligible participants were between the ages 18-40 and did not have a history of known neurological injury (including stroke, traumatic brain injury, concussion), cognitive impairment, neurodegenerative disorders, migraine headache diagnosis, seizure disorder, blindness, dysautonomia/postural orthostatic tachycardia syndrome, and history of eye surgery/amblyopia/strabismus or other congenital eye disorders that could alter pupil response before measurement. This age range of 18 to 40 years was chosen because differences in PLR differ between pediatric and adult cohorts.^{19,23} Also, pupil sizes tend to decrease after the fourth decade of life.^{19,23} Individuals who could not provide accurate measurements due to repeated blinking throughout data collection also were excluded. All subjects signed an institutional-approved informed consent form prior to participation.



Figure 1. Procedure for measuring the pupillary light reflex variables.

PROCEDURES

Procedures were developed in accordance with the Quality Appraisal of Diagnostic Reliability (QAREL) Checklist (Appendix A).^{24,25} Both the investigator and participant were seated in identical 18" tall chairs across from one another at a standardized table in an environment with fluorescent lighting. We asked subjects to focus their non-measured eye on a fixed point located 2 meters away from them to avoid accommodation of the eye being measured. The pupillometer (NeuroOptics PLR-3000, Irvine, CA), which operates in a monocular manner, was then placed against the eye (Figure 1). We used settings that were identical to those described by Asakawa et al.²¹ Settings included a positive pulse stimulus, light stimulus pulse intensity of 10 uW, and background intensity of 0 uW. The measurement duration was 5.01 s, the pulse duration was 0.80 s, and the pulse onset was immediate (0 s) to stimulate the PLR. Subjects remained as still as possible and refrained from blinking during the 5-s measurement period. The investigator recorded all PLR measurements (initial pupil diameter [INITIAL], end pupil diameter [END], % change [DELTA], constriction latency [LATENCY], average constriction velocity [ACV], maximum constriction velocity [MCV], average dilation velocity [ADV], and T75). The subjects rested between 30 seconds and one minute before the investigator measured the other eye. The investigators took three trials for the right eye and three trials for the left eye. Subjects rested one to two minutes before a second investigator repeated the same measurements. Raters examined subjects and recorded values independent of each other. The average of the three trials for each eye was used to determine inter-rater reliability; individual trials were used to determine inter-trial reliability.

STATISTICAL ANALYSIS

All analyses were conducted using IBM SPSS Statistics for Windows, Version 28 (IBM Corp, Armonk, NY, USA) with the level of significance established at the 0.05 level. Means, standard deviations, and 95% confidence intervals were calculated for all dependent measures.

Separate independent *t*-tests were used to compare group differences between rater 1 and rater 2. Separate ICC [2,3] and standard error of measurement (SEM) were used to determine inter-rater reliability and measurement precision.²⁶ The minimal detectable change (MDC) was also calculated for inter-rater reliability to determine each measure's responsiveness.²⁷ The MDC represents the minimal amount of change exceeding the SEM, and represents a change beyond measurement error.²⁷ Separate ICC [3,1] and SEM were used to determine inter-trial reliability and measurement precision for Rater 1 and Rater 2. ICC values ≤ 0.5 were indicative of poor, between 0.5 and 0.75 were indicative of moderate, between 0.75 and 0.9 were indicative of good, and ≥ 0.90 were indicative of excellent reliability.²⁸

Bland-Altman plots were used to determine the similarity between each measure. For this purpose, the difference (bias) between raters and the mean score (magnitude) for the raters were plotted to provide important information regarding bias.²⁹ Between-rater score differences that were scattered (i.e., no tendency for a score to be higher or lower) were considered unbiased. The plots also assessed for bias associated with the magnitude of a score. Bias would occur when the between-rater score differences were associated with an increase in the score magnitude.²⁹

RESULTS

No significant differences ($p > 0.05$) existed between any of the measures taken by Rater 1 and Rater 2 (Table 1). For inter-rater reliability, ICC [2,3] exceeded 0.70 for all measures except for T75 (Table 2). Four of the eight measures had excellent reliability for each eye as evidenced by ICCs exceeding 0.90 (INITIAL, END, ACV, and MCV). A similar pattern of values existed (Table 3) for inter-trial reliability.

Except for T75, the Bland-Altman plots showed a random pattern between the difference and mean for each measure (Figures 2 - 4). These plots also did not show a pattern of differences increasing or decreasing as the score magnitude (mean) increased. These factors taken together suggested no bias for these measures. For T75, the Bland-Altman plots appeared less scattered, and many differences appeared to increase with greater score magnitude. This finding suggested bias for T75, especially for the right eye.

DISCUSSION

The current study was the first to examine inter-rater reliability for the eight PLR parameters obtained using the NO3000. Except for the T75, inter-trial and inter-trial reliability was moderate to excellent for all measures. ICCs for T75 were poor to moderate and did not meet the minimum

Table 1. Means \pm standard deviations and (95% confidence intervals) for all dependent measures (n=48).

| Measure | Left Eye | | | Right Eye | | |
|-------------------------------------|--------------------------------|--------------------------------|----------|--------------------------------|--------------------------------|----------|
| | Rater 1 | Rater 2 | p-value* | Rater 1 | Rater 2 | p-value* |
| Initial pupil size, mm | 5.5 \pm 0.7 (5.3, 5.7) | 5.4 \pm 0.7 (5.2, 5.6) | 0.76 | 5.3 \pm 0.7 (5.1, 5.5) | 5.3 \pm 0.7 (5.1, 5.4) | 0.67 |
| End pupil size, mm | 4.0 \pm 0.6 (3.8, 4.2) | 4.0 \pm 0.6 (3.8, 4.2) | 0.84 | 3.9 \pm 0.6 (3.7, 4.1) | 3.9 \pm 0.6 (3.7, 4.1) | 0.95 |
| Change in pupil size, % | 26.6 \pm 5.2 (25.1, 28.1) | 26.4 \pm 4.5 (25.1, 27.7) | 0.84 | 26.6 \pm 4.7 (25.3, 28.0) | 25.5 \pm 4.6 (24.2, 26.8) | 0.22 |
| Latency, ms | 0.2 \pm 0.0 (0.2, 0.3) | 0.2 \pm 0.0 (0.2, 0.3) | 0.28 | 0.2 \pm 0.0 (0.2, 0.3) | 0.2 \pm 0.0 (0.2, 0.3) | 0.39 |
| Average constriction velocity, mm/s | -3.1 \pm 0.6 (-3.2, -2.9) | -3.2 \pm 0.6 (-3.4, -3.0) | 0.25 | -3.0 \pm 0.5 (-3.2, -2.9) | -3.1 \pm 0.6 (-3.3, -3.0) | 0.46 |
| Maximum constriction velocity, mm/s | -4.3 \pm 0.7 (-4.5, -4.0) | -4.4 \pm 0.7 (-4.6, -4.2) | 0.45 | -4.3 \pm 0.7 (-4.4, -4.1) | -4.3 \pm 0.7 (-4.5, -4.0) | 0.99 |
| Average dilation velocity, mm/s | 1.2 \pm 0.3 (1.1, 1.3) | 1.1 \pm 0.3 (1.0, 1.2) | 0.18 | 1.2 \pm 0.3 (1.1, 1.3) | 1.1 \pm 0.3 (1.0, 1.2) | 0.07 |
| 75% recovery time, s | 1.7 \pm 0.7 (1.4, 1.9) | 1.5 \pm 0.5 (1.4, 1.7) | 0.38 | 1.7 \pm 0.7 (1.4, 1.9) | 1.5 \pm 0.4 (1.4, 1.6) | 0.67 |

* Means compared using an independent *t*-test

Table 2. Summary of intraclass correlation coefficients (ICC), standard error of measurement (SEM), and minimal detectable change (MDC) for inter-rater reliability (n=48).

| Measure | Left Eye | | | Right Eye | | |
|-------------------------------------|-----------|------|------|-----------|------|------|
| | ICC [2,3] | SEM | MDC | ICC [2,3] | SEM | MDC |
| Initial pupil size, mm | 0.95 | 0.16 | 0.44 | 0.95 | 0.15 | 0.42 |
| End pupil size, mm | 0.96 | 0.12 | 0.33 | 0.94 | 0.14 | 0.39 |
| Change in pupil size, % | 0.89 | 3.09 | 8.57 | 0.88 | 3.07 | 8.51 |
| Latency, ms | 0.88 | 0.02 | 0.06 | 0.72 | 0.02 | 0.06 |
| Average constriction velocity, mm/s | 0.90 | 0.36 | 1.01 | 0.90 | 0.32 | 0.89 |
| Maximum constriction velocity, mm/s | 0.91 | 0.43 | 1.19 | 0.92 | 0.38 | 1.05 |
| Average dilation velocity, mm/s | 0.89 | 0.18 | 0.50 | 0.86 | 0.19 | 0.53 |
| 75% recovery time, s | 0.67 | 0.36 | 1.01 | 0.64 | 0.28 | 0.78 |

acceptable ICC of 0.7, suggesting that T75 may not be a useful biomarker.

INTER-RATER AND INTER-TRIAL RELIABILITY OF THE NO3000

Moderate to excellent inter-rater ICCs existed when measuring seven of the eight PLR measures using the NO3000, supporting its robustness as an automated pupillometer. Equally important was acceptable inter-trial reliability, the ability for a user to repeat measures and obtain consistent results^{12,21} Most inter-trial ICCs were good to excellent and agreed with Asakawa et al.,²¹ who used the NO3000 to examine inter-trial reliability. McKay et al.³⁰ compared measures from the NO3000 to BrightLamp, a pupillometer app, and also found strong measurement reproducibility for the NO3000. Findings from the current study further support the reliability of the NO3000.^{21,30} Master et al.¹⁹ have used PLR as a biomarker for identifying sport-related concus-

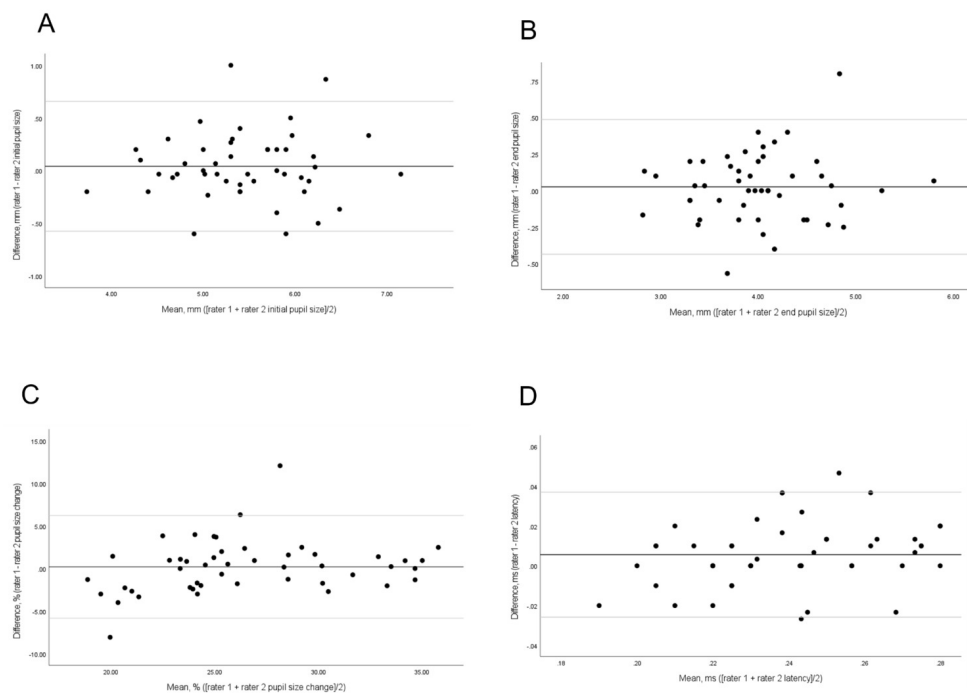
sions in adolescents; having a device with acceptable reliability is critical for clinical decision-making.³¹

INCONSISTENCIES IN THE INTER-RATER AND INTER-TRIAL RELIABILITY OF T75

The current findings suggest that both the inter-rater and inter-trial reliability of T75 were poor to moderate.³² Unacceptable T75 reliability may have resulted from measurement precision. To obtain T75, subjects must keep their eyes still throughout the entire measurement period. Researchers who have examined pediatric populations have reported sources of error from movement¹⁴ and shorter stimulus durations.³³ In studies analyzing T75 using the PLR-2000 or PLR-3000 models, the duration of the stimulus was 154 ms or 800 ms.^{14,19,34-37} This variation in stimulus duration across studies may represent a source of error contributing to unacceptable reliability.

Table 3. Summary of intraclass correlation coefficients (ICC) and standard error of measurement (SEM) for inter-trial reliability (n=48).

| Measure | Left Eye | | | | Right Eye | | | |
|-------------------------------------|--------------|------|--------------|------|--------------|------|--------------|------|
| | Rater 1 | | Rater 2 | | Rater 1 | | Rater 2 | |
| | ICC [3,1] | SEM | ICC [3,1] | SEM | ICC [3,1] | SEM | ICC [3,1] | SEM |
| Initial pupil size, mm | 0.97 | 0.38 | 0.96 | 0.43 | 0.94 | 0.51 | 0.96 | 0.39 |
| End pupil size, mm | 0.97 | 0.35 | 0.96 | 0.37 | 0.94 | 0.41 | 0.98 | 0.26 |
| Change in pupil size, % | 0.90 | 4.83 | 0.80 | 7.21 | 0.92 | 4.12 | 0.92 | 3.85 |
| Latency, ms | 0.80 | 0.04 | 0.84 | 0.03 | 0.77 | 0.04 | 0.76 | 0.03 |
| Average constriction velocity, mm/s | 0.88 | 0.61 | 0.93 | 0.47 | 0.81 | 0.65 | 0.84 | 0.66 |
| Maximum constriction velocity, mm/s | 0.92 | 0.63 | 0.92 | 0.62 | 0.86 | 0.78 | 0.90 | 0.69 |
| Average dilation velocity, mm/s | 0.88 | 0.30 | 0.92 | 0.22 | 0.89 | 0.26 | 0.91 | 0.24 |
| 75% recovery time, s | 0.62 | 0.87 | 0.41 | 1.04 | 0.54 | 1.02 | 0.65 | 0.74 |

**Figure 2. Bland-Altman plots comparing differences in measures for the left eye between rater 1 and rater 2 for initial pupil size (A), end pupil size (B), pupil size change (C), and latency (D). The solid black line represents the mean of the difference between raters. The gray-dashed lines represent the mean difference \pm 2 standard deviations.**

The current findings coincided with Asakawa et al.,²¹ who also used the NO3000. Asakawa et al. reported poor inter-trial T75 reliability and suggested that specified device settings be used to obtain this parameter.²¹ They used a 180- μ watt/cm² stimulus with an 800-ms duration that was considerably higher compared to our 30-ms duration.²¹ Asakawa et al. concluded that poor reliability resulted from a lack of optimal settings to accommodate the time required for the eye to reach 75% of its baseline size. The time between trials also may need to be lengthened and standardized to give the eye adequate time to recover before be-

ing re-stimulated. Yoo et al.³⁷ recorded for a 5 s duration after initiating a 180 μ watts/cm² stimulus for 185 ms and found significant differences in T75 and pupil diameters between healthy individuals and those with Horner Syndrome. These settings differed from the settings used in the current research of 180 μ watts/cm² for 30 ms and suggested the light stimulus settings, particularly the duration, be increased to obtain consistent T75 data. Others^{14,19} that used T75 analysis in populations with concussions used pupilometer settings with a 180- μ watt/cm² but a 154 ms stimulus duration. Findings from these studies were more con-

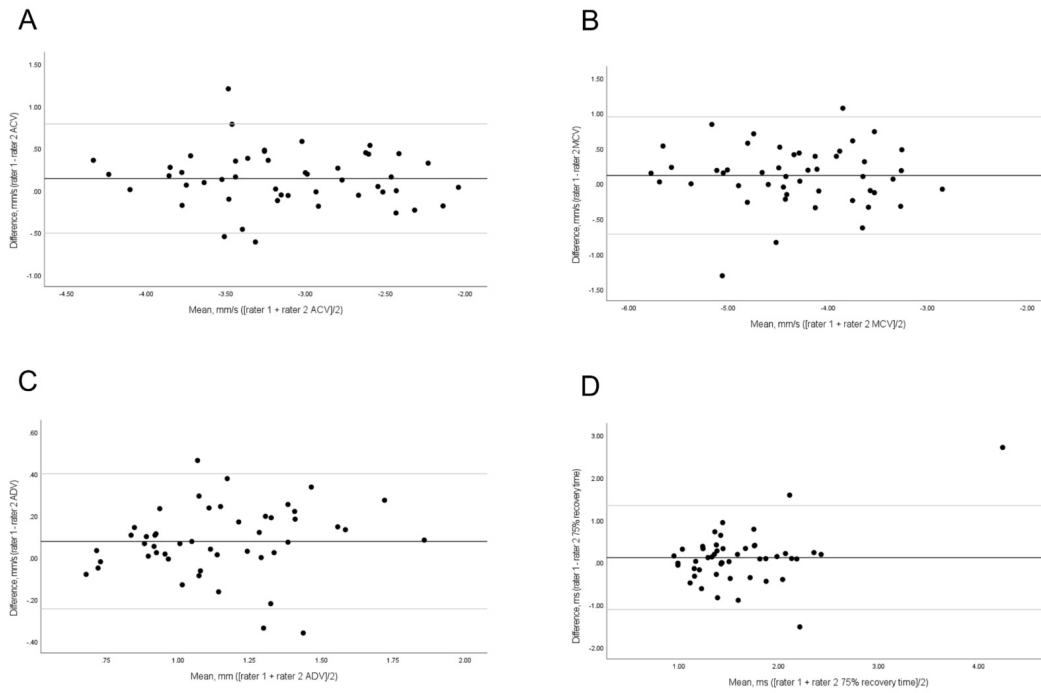


Figure 3. Bland-Altman plots comparing differences in measures for the left eye between rater 1 and rater 2 for average constriction velocity (A), maximum constriction velocity (B), average dilation velocity (C), and 75% time to recovery (D). The solid black line represents the mean of the difference between raters. The gray-dashed lines represent the mean difference \pm 2 standard deviations.

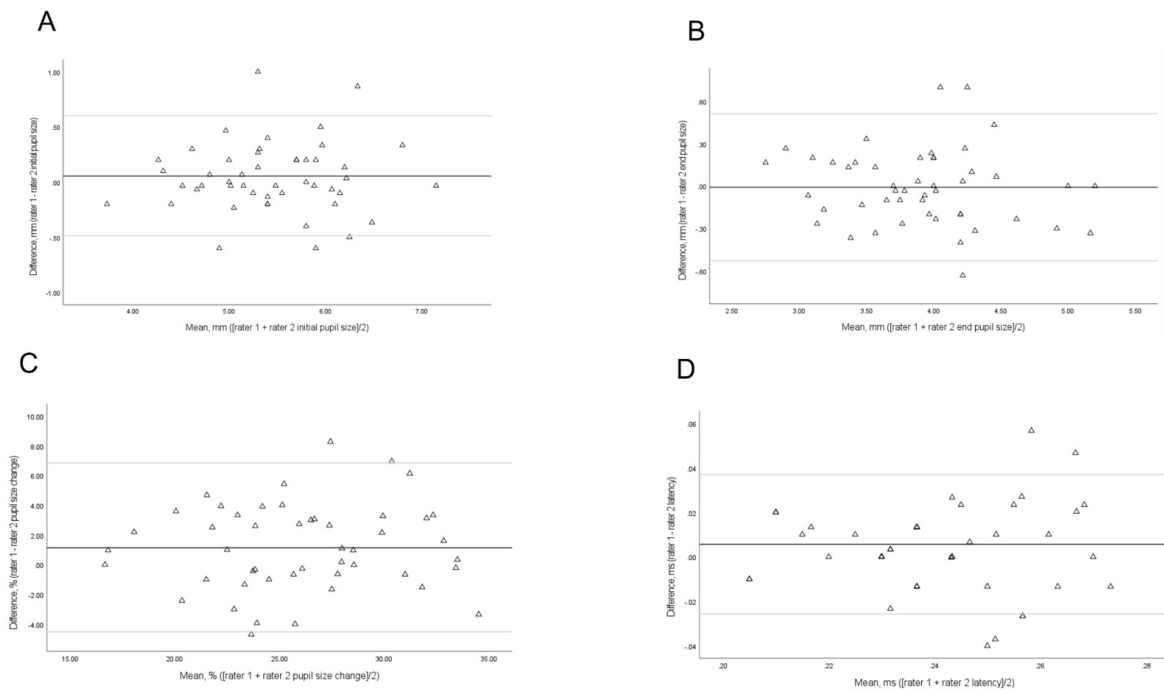


Figure 4. Bland-Altman plots comparing differences in measures for the right eye between rater 1 and rater 2 for initial pupil size (A), end pupil size (B), pupil size change (C), and latency (D). The solid black line represents the mean of the difference between raters. The gray-dashed lines represent the mean difference \pm 2 standard deviations.

sistent, which suggests the importance of a longer duration range to obtain reliable T75 data.^{14,19} Future investigators should pay special attention to the light stimulus intensity

and duration when obtaining PLR parameters and which testing conditions are needed to optimize data collection.

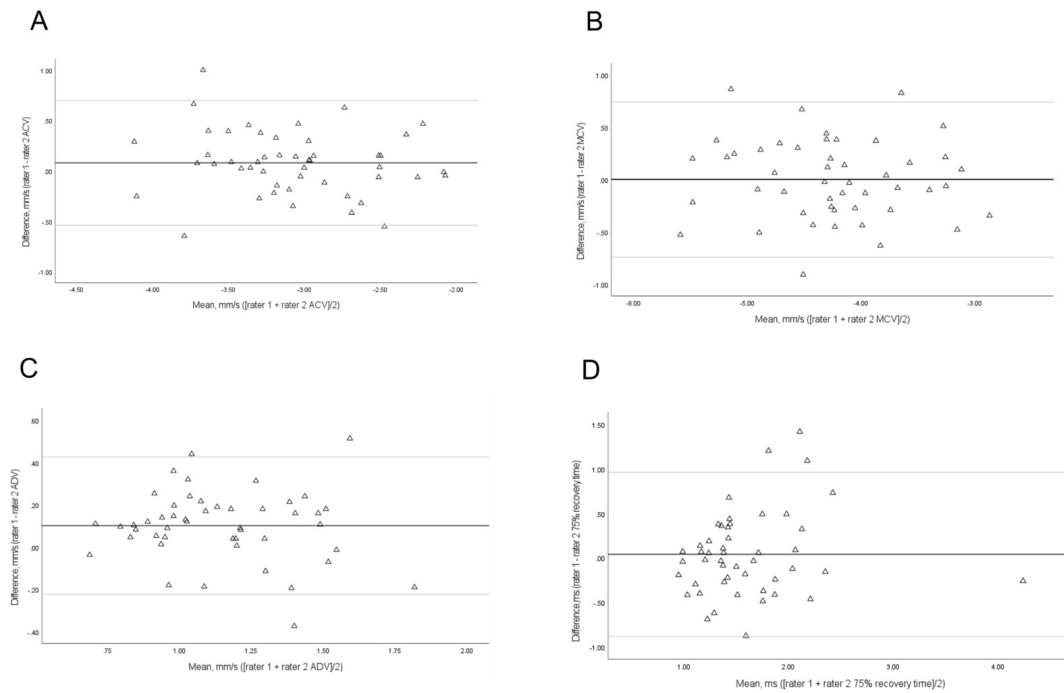


Figure 5. Bland-Altman plots comparing differences in measures for the right eye between rater 1 and rater 2 for average constriction velocity (A), maximum constriction velocity (B), average dilation velocity (C), and 75% time to recovery (D). The solid black line represents the mean of the difference between raters. The gray-dashed lines represent the mean difference \pm 2 standard deviations.

CLINICAL IMPLICATIONS

Measurement reliability is critical to enhance clinical decision-making.³¹ It supports that a change in a parameter represents a “true” change in the behavior. Clinicians also can use the MDC values (Table 2) to determine if changes in a measure exceed the inherent measurement variability, thus representing a “true” change.²⁷ Other previous pupillometer models from NeuroOptics have established inter-trial and inter-device reliability, supporting their use in the critical care field for the evaluation of traumatic brain injury.¹⁷ Findings from the current study generally support the use of the NO3000 to assess PLR in screening settings and research applications. The NO3000 is useful in detecting concussions because there are known changes to the pupillary light response following trauma.⁴ Pupillometers can increase detection, especially when clinical symptoms may be lacking, and mitigate human error.^{3,4} With their user-friendly and portable designs, settings beyond research labs such as sports medicine or physical therapy clinics can use this device to monitor recovery progress.^{3,4} However, caution is required when measuring T75 due to more sources of error. Future investigations should determine the optimal settings for this measure.

LIMITATIONS

This study has limitations. Only one setting of the light stimulus and recording period was used during data collection. Having conducted trials using different light stimulus intensities and durations could have elicited ranges that

accommodate and improve the reproducibility of T75. Eye dominance in our participants was not determined, thus, the authors are unable to explain the discrepancies in T75 observed in the right eye only rather than both eyes. Only the most recent version of the model was used, making the current findings only generalizable to the NO3000. A final limitation was the use of healthy subjects, which has limited generalization to clinical populations.

CONCLUSION

The inter-rater and inter-trial reliability of the NO3000 was established. All parameters, except T75, exhibited good to excellent inter-rater and inter-trial reliability. T75 had moderate inter-rater and inter-trial reliability, which likely reflected inherent challenges when obtaining this measure. The NO3000 can be used in future pupillometry studies focused on measuring static and dynamic PLR parameters, but attention and rationale regarding the stimulus settings and environment are needed to minimize measurement error. Further investigation is needed to examine if other pupillometers can reliably measure T75 using different light stimulus intensities and durations.

CONFLICTS OF INTEREST

The authors report no conflicts of interest.

Submitted: April 18, 2024 CST. Accepted: November 22, 2024

CST. Published: February 01, 2025 CST.

© The Author(s)



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

REFERENCES

1. Couret D, Boumaza D, Grisotto C, et al. Reliability of standard pupillometry practice in neurocritical care: an observational, double-blinded study. *Crit Care*. 2016;20:99. doi:[10.1186/s13054-016-1239-z](https://doi.org/10.1186/s13054-016-1239-z)
2. Taylor WR, Chen JW, Meltzer H, et al. Quantitative pupillometry, a new technology: normative data and preliminary observations in patients with acute head injury. Technical note. *J Neurosurg*. 2003;98(1):205-213. doi:[10.3171/jns.2003.98.1.0205](https://doi.org/10.3171/jns.2003.98.1.0205)
3. Boulter JH, Shields MM, Meister MR, Murtha G, Curry BP, Dengler BA. The expanding role of quantitative pupillometry in the evaluation and management of traumatic brain injury. *Front Neurol*. 2021;12:685313. doi:[10.3389/fneur.2021.685313](https://doi.org/10.3389/fneur.2021.685313)
4. Vrettou CS, Fragkou PC, Mallios I, et al. The role of automated infrared pupillometry in traumatic brain injury: a narrative review. *J Clin Med*. 2024;13(2). doi:[10.3390/jcm13020614](https://doi.org/10.3390/jcm13020614)
5. Olson DM, Stutzman S, Saju C, Wilson M, Zhao W, Aiyagari V. Interrater reliability of pupillary assessments. *Neurocrit Care*. 2016;24(2):251-257. doi:[10.1007/s12028-015-0182-1](https://doi.org/10.1007/s12028-015-0182-1)
6. Hall CA, Chilcott RP. Eyeing up the future of the pupillary light reflex in neurodiagnostics. *Diagnostics*. 2018;8(1). doi:[10.3390/diagnostics8010019](https://doi.org/10.3390/diagnostics8010019)
7. Mathot S. Pupillometry: psychology, physiology, and function. *J Cogn*. 2018;1(1):16. doi:[10.5334/joc.18](https://doi.org/10.5334/joc.18)
8. Ciuffreda KJ, Joshi NR, Truong JQ. Understanding the effects of mild traumatic brain injury on the pupillary light reflex. *Concussion*. 2017;2(3):CNC36. doi:[10.2217/cnc-2016-0029](https://doi.org/10.2217/cnc-2016-0029)
9. Troiani V. The future of quantitative pupillometry in health and disease. *Clin Auton Res*. 2020;30(1):11-12. doi:[10.1007/s10286-019-00655-3](https://doi.org/10.1007/s10286-019-00655-3)
10. Park JG, Moon CT, Park DS, Song SW. Clinical utility of an automated pupillometer in patients with acute brain lesion. *J Korean Neurosurg Soc*. 2015;58(4):363-367. doi:[10.3340/jkns.2015.58.4.363](https://doi.org/10.3340/jkns.2015.58.4.363)
11. Chen JW, Gombart ZJ, Rogers S, Gardiner SK, Cecil S, Bullock RM. Pupillary reactivity as an early indicator of increased intracranial pressure: The introduction of the Neurological Pupil index. *Surg Neurol Int*. 2011;2:82. doi:[10.4103/2152-7806.82248](https://doi.org/10.4103/2152-7806.82248)
12. Nyholm B, Obling L, Hassager C, et al. Superior reproducibility and repeatability in automated quantitative pupillometry compared to standard manual assessment, and quantitative pupillary response parameters present high reliability in critically ill cardiac patients. *PLoS One*. 2022;17(7):e0272303. doi:[10.1371/journal.pone.0272303](https://doi.org/10.1371/journal.pone.0272303)
13. Hershaw JN, Ettenhofer ML. Insights into cognitive pupillometry: Evaluation of the utility of pupillary metrics for assessing cognitive load in normative and clinical samples. *Int J Psychophysiol*. 2018;134:62-78. doi:[10.1016/j.ijpsycho.2018.10.008](https://doi.org/10.1016/j.ijpsycho.2018.10.008)
14. Hsu J, Stec M, Ranaivo HR, Srdanovic N, Kurup SP. Concussion alters dynamic pupillary light responses in children. *J Child Neurol*. 2021;36(3):195-202. doi:[10.1177/0883073820964040](https://doi.org/10.1177/0883073820964040)
15. Jahns FP, Miroz JP, Messerer M, et al. Quantitative pupillometry for the monitoring of intracranial hypertension in patients with severe traumatic brain injury. *Crit Care*. 2019;23(1):155. doi:[10.1186/s13054-019-2436-3](https://doi.org/10.1186/s13054-019-2436-3)
16. Meeker M, Du R, Bacchetti P, et al. Pupil examination: validity and clinical utility of an automated pupillometer. *J Neurosci Nurs*. 2005;37(1):34-40. doi:[10.1097/01376517-200502000-00006](https://doi.org/10.1097/01376517-200502000-00006)
17. Stutzman S, Iype P, Marshall J, et al. Inter-device reliability of the NPi-200 and NPi-300 pupillometers. *J Clin Neurosci*. 2022;100:180-183. doi:[10.1016/j.jocn.2022.04.023](https://doi.org/10.1016/j.jocn.2022.04.023)
18. Patwari PP, Stewart TM, Rand CM, et al. Pupillometry in congenital central hypoventilation syndrome (CCHS): quantitative evidence of autonomic nervous system dysregulation. *Pediatr Res*. 2012;71(3):280-285. doi:[10.1038/pr.2011.38](https://doi.org/10.1038/pr.2011.38)
19. Master CL, Podolak OE, Ciuffreda KJ, et al. Utility of pupillary light reflex metrics as a physiologic biomarker for adolescent sport-related concussion. *JAMA Ophthalmol*. 2020;138(11):1135-1141. doi:[10.1001/jamaophthalmol.2020.3466](https://doi.org/10.1001/jamaophthalmol.2020.3466)
20. Hall CA, Chilcott RP. Eyeing up the future of the pupillary light reflex in neurodiagnostics. *Diagnostics (Basel)*. 2018;8(1):19. doi:[10.3390/diagnostics8010019](https://doi.org/10.3390/diagnostics8010019)
21. Asakawa KI, H. Reproducibility and normative values of the parameters of a new hand-held digital pupillometer. *J Clin Exp Ophthalmol*. 2017;8:654.

22. Arifin WN. A web-based sample size calculator for reliability studies. *Educ Med J*. 2018;10(3):67-76. doi:[10.21315/eimj2018.10.3.8](https://doi.org/10.21315/eimj2018.10.3.8)
23. Rickmann A, Waizel M, Kazerounian S, Szurman P, Wilhelm H, Boden KT. Digital pupillometry in normal subjects. *Neuroophthalmology*. 2017;41(1):12-18. doi:[10.1080/01658107.2016.1226345](https://doi.org/10.1080/01658107.2016.1226345)
24. EQUATOR Network. Enhancing the QUALity and Transparency Of health Research. Accessed August 15, 2023. <https://www.equator-network.org/reporting-guidelines/recommendations-for-reporting-the-results-of-studies-of-instrument-and-scale-development-and-testing/>
25. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. *J Adv Nurs*. 2014;70(9):1970-1979. doi:[10.1111/jan.12402](https://doi.org/10.1111/jan.12402)
26. Denegar C, Ball D. Assessing reliability and precision of measurement: an introduction to intraclass correlation and standard error of measurement. *J Sports Rehabil*. 1993;2:35-42. doi:[10.1123/jsr.2.1.35](https://doi.org/10.1123/jsr.2.1.35)
27. Jewell DV. Appraising the evidence. In: *Guide to Evidence-Based Physical Therapist Practice*. 5th ed. Jones & Bartlett Learning; 2023:277-290.
28. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163. doi:[10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)
29. Hergott CG, Bolgla LA, Waller JL, et al. TUG-10: a modification of the timed up and go test for aerobic assessment in older adults. *Cardiopulm Phys Ther J*. 2022;33(4):157-165. doi:[10.1097/CPT.000000000000202](https://doi.org/10.1097/CPT.000000000000202)
30. McKay RE, Kohn MA, Schwartz ES, Larson MD. Evaluation of two portable pupillometers to assess clinical utility. *Concussion*. 2020;5(4):CNC82. doi:[10.2217/cnc-2020-0016](https://doi.org/10.2217/cnc-2020-0016)
31. Portney LG. Concepts of measurement reliability. In: *Foundations of Clinical Research. Applications to Evidence-Based Practice*. 4th ed. FA Davis; 2020:115-125.
32. Portney LG, Gross KD. Measurement revisited: reliability and validity statistics. In: *Foundations of Clinical Research. Applications to Evidence-Based Practice*. 4th ed. FA Davis; 2020:486-508.
33. Winston M, Zhou A, Rand CM, et al. Pupillometry measures of autonomic nervous system regulation with advancing age in a healthy pediatric cohort. *Clin Auton Res*. 2020;30(1):43-51. doi:[10.1007/s10286-019-00639-3](https://doi.org/10.1007/s10286-019-00639-3)
34. Bista Karki S, Coppell KJ, Mitchell LV, Ogbuehi KC. Dynamic pupillometry in type 2 diabetes: pupillary autonomic dysfunction and the severity of diabetic retinopathy. *Clin Ophthalmol*. 2020;14:3923-3930. doi:[10.2147/OPTH.S279872](https://doi.org/10.2147/OPTH.S279872)
35. Lynch GTF, James SM, Cardon TA, McPherson SM. Sensitivity and specificity of pupillary light reflex measures for ASD using monocular pupillometry. *Neurol Sci*. 2022;43(7):4537-4545. doi:[10.1007/s10072-022-05976-2](https://doi.org/10.1007/s10072-022-05976-2)
36. Papangelou A, Boorman DW, Sharifpour M, Patel HP, Cassim T, Garcia PS. Associations of an eye-tracking task and pupillary metrics with age and ASA physical status score in a preoperative cohort. *J Clin Monit Comput*. 2023;37(3):795-803. doi:[10.1007/s10877-023-00974-x](https://doi.org/10.1007/s10877-023-00974-x)
37. Yoo YJ, Yang HK, Hwang JM. Efficacy of digital pupillometry for diagnosis of Horner syndrome. *PLoS One*. 2017;12(6):e0178361. doi:[10.1371/journal.pone.0178361](https://doi.org/10.1371/journal.pone.0178361)

SUPPLEMENTARY MATERIALS

Appendix A

Download: <https://ijspt.scholasticahq.com/article/128047-assessing-the-inter-rater-and-inter-trial-reliability-of-the-neurooptics-pupillary-light-response-3000-pupillometer/attachment/260638.docx>
