USAARL Report No. 2018-07

Evaluation of the Commercial, Off-the-Shelf (COTS) King-Devick Eye Tracking System

By Leonard A. Temme[1], Paul St. Onge[1,2], Aaron McAtee[1,3], Kevin J. O'Brien[1,3], Brigid K. Byrd[1,3]

[1]U.S. Army Aeromedical Research Laboratory
[2]Laulima Government Solutions, LLC
[3]Oak Ridge Institute for Science and Education

United States Army Aeromedical Research Laboratory

Visual Protection and Performance Division

March 2018

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
19-03-2017	Final	

4. TITLE AND SUBTITLE
Evaluation of the Commercial, Off-the-Shelf (COTS) King-Devick Eye Tracking System

5a. CONTRACT NUMBER N/A
5b. GRANT NUMBER N/A
5c. PROGRAM ELEMENT NUMBER N/A

6. AUTHOR(S)
Temme, Leonard A.; St. Onge, Paul; McAtee, Aaron; O'Brien, Kevin J.; Byrd, Brigid K.

5d. PROJECT NUMBER N/A
5e. TASK NUMBER N/A
5f. WORK UNIT NUMBER N/A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL 2018-07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
N/A

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
Laulima Government Solutions LLC; Oak Ridge Institute for Science and Education

14. ABSTRACT
The identification and characterization of sensitive biomarkers for concussion are important for military medicine to guide diagnosis and return to duty decisions. The current literature suggests that the King Devick (KD) number reading test is a sensitive screener for concussion sustained during sports activities. The KD test involves timing an individual reading aloud 120 single digit numbers, without error, printed on three standard cards; timing is performed with a stopwatch. The extensive literature commonly describes the KD test as revealing the disruption of eye movements that could be due to a number of factors including concussion. Thus, the test is sensitive if not necessarily specific. The extent to which the KD test reflects eye movements remains a matter of conjecture since it is a measure of gross behavioral reading performance. Hence, it is important to objectively measure eye movements concurrently with the performance of the KD test. Recently, a commercial, off-the-shelf (COTS), integrated KD / Eye Tracking system became available.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
UNCLAS	UNCLAS	UNCLAS	SAR	43	Loraine St. Onge, PhD
					19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

Reset

REPORT DOCUMENTATION PAGE (SF298)
(Continuation Sheet)

14. Abstract
Because of the potential importance of the contributions of this system to concussion research, two copies of the system were purchased and evaluated before they were deployed for research. The evaluation consisted of two studies; one study with 20 volunteers measured the comparability of the two systems, while the other study with 5 volunteers assessed the stability and repeatability of the measurements over 5 successive days. The results showed that several response parameters of the systems lacked face validity, calling into question the validity of the eye tracking data. Thus, we have not used these systems for further research.

Foreword

The identification and characterization of sensitive biomarkers for concussion are important for military medicine to guide diagnosis and return to duty decisions. The current literature suggests that the King Devick (KD) number reading test is a sensitive screener for concussion sustained during sports activities. The KD test involves timing an individual reading aloud 120 single digit numbers, without error, printed on three standard cards; timing is performed with a stopwatch. The extensive literature commonly describes the KD test as revealing the disruption of eye movements that could be due to a number of factors including concussion. Thus, the test is sensitive if not necessarily specific. The extent to which the KD test reflects eye movements remains a matter of conjecture since it is a measure of gross behavioral reading performance. Hence, it is important to objectively measure eye movements concurrently with the performance of the KD test. Recently, a commercial, off-the-shelf (COTS), integrated KD / Eye Tracking system became available. Because of the potential importance of the contributions of this system to concussion research, two copies of the system were purchased and evaluated before they were deployed for research. The evaluation consisted of two studies; one study with 20 volunteers measured the comparability of the two systems, while the other study with 5 volunteers assessed the stability and repeatability of the measurements over 5 successive days. The results showed that several response parameters of the systems lacked face validity, calling into question the validity of the eye tracking data. Thus, we have not used these systems for further research.

This page is intentionally blank.

Table of Contents

Page

Introduction 1
Methods 3
    Volunteers 3
    Equipment 4
    Procedures 5
    Design 6
    Data Analysis 7
Results 7
    Determining Limits of Agreement 11
    Calculation of the Coefficient of Agreement 12
Discussion 14
References 17
Appendix A. Bland-Altman plots for the response parameters 22
Appendix B. Subject-specific agreement charts 29

List of Figures

1. The two ET/KD systems. 4
2. The four cards comprising the KD test. 5
3. Typical eye tracking plots generated by the ET/KD. 8
4. Difference vs mean for Total Reading Time for trial 1 of the first experiment. 12
5. Subject-specific agreement for total response time for the first experiment. 13

List of Tables

1. The 28 eye movement output parameters generated by the ET/KD for each test card. 6
2. Selected mean measurement values by system and trial number for the first experiment. 9
3. Average values for the 1-day and 5-day experiments. 10
4. Measures of agreement for trial 1 of the first experiment. 12
5. Measures of agreements of mean differences and COA between the two ET/KD systems for the output parameters for the first experiment. 14

This page is intentionally blank.

Introduction

The increased world-wide recognition of the importance of sports-related concussion at essentially all levels of play, from elementary school through professional leagues, is matched by the increasing interest in the identification of practicable, sensitive, sideline concussion screening methods to guide the management of concussion during sporting activities and return to play decisions (Broglio et al., 2017; Khodaee, Currie, Asif, & Comstock, 2016; Leong, Balcer, Galetta, Liu, & Master, 2014; Marshall, Guskiewicz, Shankar, McCrea, & Cantu, 2015; Ventura, Balcer, Galetta, & Rucker, 2016). The Department of Defense (DoD) has a corresponding interest in methods to guide medical fitness and return to duty decisions. The DoD reports a total of 361,092 cases of traumatic brain injury between 2000 and 2016 with concussion accounting for 297,478 of these cases (Defense and Veterans Brain Injury Center (DVBIC), DoD worldwide numbers for TBI, http://dvbic.dcoe.mil/files/tbi-numbers/DoD-TBI-Worldwide-Totals_2000-2016_Feb-17-2017_v1.0_2017-04-06.pdf. These numbers underscore the importance of concussion detection for the military. The assessment of concussion in a military context is particularly challenging because the trauma often involves a combat situation, which may include blast, the "chaos of combat," and the vital concerns of securing oneself, others, and equipment, while ensuring mission success. These factors add to the difficulty of military medical and psychological evaluations, which may occur months and even years after the event (Davenport, 2016). The well-known comorbidity of concussion with post-traumatic stress syndrome, depression, and altered mental states further complicates the evaluation of concussion in the military context (Hoge et al., 2008; Kennedy et al., 2007; Maguen, Lau, Madden, & Seal, 2012; Schmid & Tortella, 2012; Seal et al., 2016). While blast exposure is a combat-related cause of concussion, military personnel can sustain concussion from a wide variety of activities, including those that affect the civilian sector (Regasa, Thomas, Gill, Marion, & Ivins, 2016). In combat, as well as in training scenarios, concussion can compromise performance and reduce situational awareness, which can quickly translate into catastrophic mission failure and loss of life.

Return to duty decisions should be guided by accurate and practical metrics, particularly since concussion can exacerbate the deleterious effects of subsequent concussions (Barker et al., 2017; Iverson, Echemendia, Lamarre, Brooks, & Gaetz, 2012; Ventura et al., 2016). Additionally, repeated concussions may be associated with the development of long-term neurodegeneration (see: http://dvbic.dcoe.mil/article/cumulative-concussions) (Ventura et al., 2016). Because of these and other complications, the military has recognized the need for clear, objective, sensitive, and specific tests and biomarkers for concussion (Marion, Curley, Schwab, & Hicks, 2011; Schmid & Tortella, 2012).

Recent reports of retrospective analyses of the medical records show that as many as 70 to 85% of concussed patients report visual dysfunctions, complaints, or impairments even when their eyes are healthy and normal upon medical examination (Brahm et al., 2009; Capo-Aponte et al., 2016; Padula, Capo-Aponte, Padula, Singman, & Jenness, 2017). In other words, following concussion, vision complaints are common in the absence of apparent or manifest ocular trauma. The list of reported visual complaints is extensive and includes visual field defects, visual neglect, visual midline shift, pupillary abnormalities, color vision deficits, photophobia, eye strain, as well as difficulties with vergence, accommodation, and reading. It should be no surprise that concussion and other traumatic brain injuries affect vision since more than 30 regions of the brain are involved with vision, seven of the twelve cranial nerves subserve vision, and more than 70% of the brain's sensory processing involves vision. Reading is among the most frequent visual difficulties following concussion (Capo-Aponte et al., 2016), and it has long been known that reading is exquisitely dependent on eye movements, which are themselves sensitive to concussion (Leigh & Zee, 2006; Liversedge, Gilchrist, & Everling, 2011).

Eye movements can be conveniently categorized as saccades, smooth pursuits, and fixation-related behaviors, which include micro-saccades to correct drift and maintain fixation (Barnes, 2011; Gilchrist, 2011; Martinez-Conde & Macknik, 2011). Saccades are characteristically rapid, ballistic, and relatively accurate sudden shifts of gaze from one area of the visual field to another. In contrast, during smooth pursuit eye movements (SPEM), gaze typically follows the continuous path of a smoothly moving target so that the image of the target remains relatively constant on the same retinal location, which is most naturally the fovea. Unlike saccades, SPEM are almost never generated in the absence of an appropriately moving stimulus (Boyer, Portron, Bevilacqua, & Lorenceau, 2017).

A recent report compared eye movements recorded in 60 blast-concussed military volunteers with eye movements recorded in 20 civilian controls (Cifu et al., 2015). Neurologic and ophthalmologic exams documented no vision problems, yet there were robust, significant, and characteristic differences in eye movements between the two groups. Specifically, the saccades of the concussed group showed larger positional errors, smaller amplitudes, slower peak velocities, slower peak accelerations, and longer durations. In general, concussed individuals tracked stepwise moving targets less accurately than the controls. These findings confirm a previous study by Heitger et al. (2009) and are consistent with the more recent report by Balaban et al. (2016) that achieved a high degree of classification (concussed vs. controls) sensitivity (89%) and specificity (97%) using measures of saccadic eye movements combined with vestibulo-ocular eye movements.

The King-Devick number reading (KDNR) test is a widely used, rapid sideline screening tool for sport-related concussion (e.g., (K. M. Galetta, Barrett, et al., 2011; K. M. Galetta, Brandes, et al., 2011; K. M. Galetta et al., 2015; M. S. Galetta et al., 2013; Leong et al., 2014; Molloy, Murphy, & Gissane, 2017; Munce et al., 2014; Rizzo, Hudson, Dai, Desai, et al., 2016). Although the literature frequently discusses the KDNR test as a direct measure of eye movements and saccades, it is more a measure of the rapid reading and verbalization of numbers (Handmaker & Waldorf, 2013). This makes the KDNR test an assessment of a chain of interdependent behaviors, some part of which is arguably dependent on saccadic eye movements and fixations crucial to reading text, even though the KDNR test does not directly reflect eye movements in isolation from other aspects of reading behavior. The KDNR test is purportedly resistant to false positives resulting from simply exercise or physical exertion associated with sports activities (Dziemianowicz et al., 2012; Leong et al., 2014). It is also sensitive to such neurological diseases such as Alzheimer's, Parkinsonism, and multiple sclerosis (Lin, Adler, et al., 2014; Lin, Rigby, et al., 2014; Moster, Wilson, Galetta, & Balcer, 2014; Rizzo, Hudson, Dai, Desai, et al., 2016), as well as to the effects of sleep deprivation (Davies, Henderson, Balcer, & Galetta, 2012) and hypoxia (Stepanek et al., 2013).

Recently, the KDNR test has been integrated with eye tracking technology to enable administration of the KDNR test simultaneously with the measurement of eye movements

(Rizzo, Hudson, Dai, Birkemeier, et al., 2016; Rizzo, Hudson, Dai, Desai, et al., 2016; Stepanek et al., 2013). Rizzo et al. (2016a,b) used such eye tracking instrumentation to record eye movements in 42 controls and 25 concussed civilians as these volunteers performed the KDNR test. While there were no differences between the two groups in saccadic velocity, duration, or amplitude, the concussed group produced a significantly larger number of saccades with significantly less precisely controlled endpoints. Furthermore, the time between successive saccades was longer for the concussed than the control group; differences that are completely consistent with the longer time the concussed group needed to perform the KDNR test. Thus, the KDNR test provides a gross measure of reading behaviors that can reflect alterations in eye movement behaviors associated with reading, and these alterations in eye movements can be associated with concussion.

Because of the need for objective concussion biomarkers, and the potential of reading-related eye movements as a concussion biomarker, the present study evaluated a recently available, commercial, off the shelf (COTS) eye tracker / KDNR test (ET/KD) system. This system is purportedly a turnkey system ready for use upon delivery. Two of these ET/KD systems were purchased to support multiple research programs assessing concussion, hypoxia, and sleep deprivation. Before deploying these new COTS ET/KD systems for extended data collection, prudence dictated the evaluation of the reliability, validity, and comparability of the two presumably interchangeable systems, particularly since they were to be deployed in parallel at different remote testing sites.

Methods

This study's protocol (No. M-10467 USAARL2015-019) was reviewed and approved by the U. S. Army Medical Research and Materiel Command Institutional Review Board (USAMRMC IRB).

Volunteers

The volunteers were a self-selected sample from the population of adults within USAARL's immediate geographic locale. All volunteers were at least 19 years of age; and as determined by screening with the Armed Forces Vision Tester, they all had normal binocular vision, and at least 20/30 near visual acuity either with or without single power corrective lenses (contacts or glasses). Volunteer exclusion criteria included the presence of an intraocular lens implant or use of bifocal, progressive, or other multi-focal corrective lenses at the time of testing. There were no exclusion criteria for this study based on gender, ethnicity, race, or other demographic characteristics. Volunteers were not financially reimbursed for their participation.

Equipment

Two copies of the same ET/KD system were ordered at the same time from the manufacturer (Figure 1). Each system consists of a laptop computer and proprietary software integrated with a third party eye tracker. The systems arrived with the software installed. Resident software functions included calibration of the system's eye tracker and volunteer's eye position and gaze, test administration, and data management. The two systems were understood to be identical and completely functional on arrival.

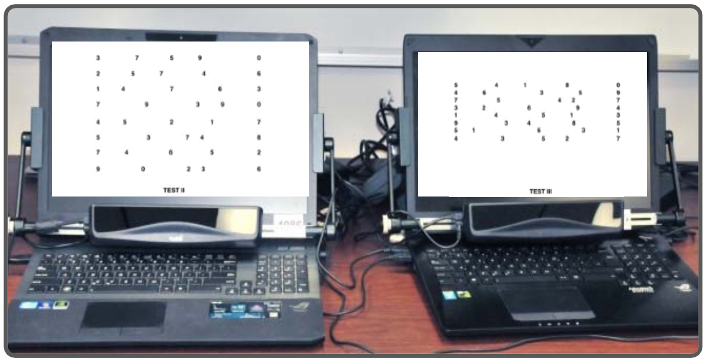


Figure 1. The two ET/KD systems. The eye tracker is located below the laptop computer's flip-up screen. The screen on the left shows Test Card (TC) 2 while the screen on the right shows TC 3.

The ET/KD system presented the four standard KD cards on the computer screen. These cards are reproduced in Figure 2. The first card is the demonstration card (upper left); it contains five rows of five unevenly spaced single-digit numbers. These numbers are connected with line segments and arrow heads to indicate the path the eye's line of sight is intended to follow as the eye scans the set of 25 numbers on the card. Reading performance on the demonstration card is not scored; the card is solely for demonstration purposes. The first test card (TC 1) also contains eight rows of five single-digit numbers. All the first digits of the eight rows are vertically aligned as are all the last digits, thus defining a left and right margin, respectively. The three inner digits of each of the eight rows are spaced irregularly. Every row includes horizontal line segments between the digits to make the intended scan path explicit. The second test card (TC 2) contains 40 single-digit numbers in essentially the same layout as TC 1: 8 rows with the left and right edges justified and 3 irregularly spaced inner digits in each row. The difference between TC 1 and TC 2 is that TC 2 does not contain any line segments to aid visual scanning along the horizontal. The third test card (TC 3) again contains 40 single-digit numbers presented in 8 rows of 5 digits each. All the first digits of the 8 rows are vertically aligned as are the last digits, establishing, respectively, a left and right edge with the 3 remaining digits irregularly spaced on each row. The difference between TC 2 and TC 3 is that the vertical spacing has been compressed on TC 3 by about 50%. The expectation is that the difficulty reading the digits increases from TC 1 to TC 2 to TC 3, and that this increased reading difficulty is reflected in reading performance. The conventional KDNR test metric is the length of time required to read the three test cards, typically reported in seconds measured with a stopwatch level of precision. Additionally, the number of errors made while reading aloud the individual numbers on each of the three TCs is recorded.

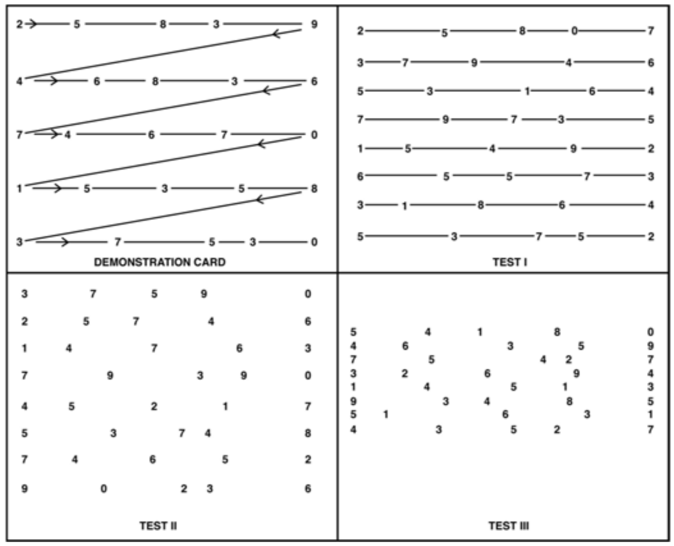


Figure 2. The four cards comprising the KD test. These same four cards are presented in the ET/KD systems. The Demonstration Card in the upper left quadrant is used only to demonstrate the test to the subject. The three remaining cards identified as TEST I, TEST II, and TEST III comprise the test itself. For all four cards, subjects read aloud each number, starting in the upper left of the card, continuing from left to right, and at the end of each line, moving down to the leftmost number on the next lower line.

Procedures

Each prospective volunteer was consented individually and in the privacy of the laboratory by one of the authors. Following the informed consent process, the volunteer's vision was screened using the Armed Forces Vision Tester to ensure that the volunteer's vision satisfied the selection criteria. Following the screening, the volunteer was seated in front of the appropriate ET/KD system, and the technician sitting beside the volunteer demonstrated the test using the second ET/KD system beside the volunteer's test system. Any additional questions the volunteer may have had as a result of the demonstration were answered.

When the volunteer was ready to begin data collection, the volunteer initiated testing with the ET/KD system, which administers the test under computer control. During testing, the test program steps through 11 screens: (1) General Instructions, (2) Calibration, (3) Calibration Results & Recalibration Option, (4) Test Instructions, (5) Demonstration Card, (6) Test Card I, (7) Test Card I Completed Screen, (8) Test Card II, (9) Test Card II Completed Screen, (10) Test Card III, and (11) Save Data Screen. The volunteer controls the progress and pacing from one screen to the next.

After calibration is complete, the Test Instructions screen presents the following text: *"The test will begin automatically after the Demonstration Screen. Please read the numbers aloud on Test Card I, starting with the number on the top left, immediately. As soon as you read the final number on each Test Card, click Next. Read aloud the numbers on each test card as quickly as you can. The Test will end when Test Card III is completed by clicking Next.*

Please Note: Do not use your finger or the mouse to assist in reading the numbers on the Test Cards."

All testing was done under a photopic ambient lighting condition with fluorescent ceiling lighting of approximately 1000 lux, which is typical for office lighting. Since the manufacturer asserted that head stabilization was not necessary with the ET/KD system, the present studies did not use a head and chin rest.

For the first experiment, 20 volunteers alternated between the two ET/KD systems on successive trials. The two ET/KD systems were set up next to each other on the table. A volunteer's participation in the first experiment, from informed consent to the completion of data collection, invariably required one test session lasting less than two hours, including the volunteer's self-paced rest breaks.

The second experiment required five volunteers to return to the laboratory on five successive days at approximately the same time of day. On each of these days, a volunteer made a pair of readings on ET/KD systems, alternating between the two systems to produce four readings a day. The first day's test session included the informed consent and lasted about an hour; the 4 subsequent sessions lasted about 30 minutes per volunteer.

In both experiments, the ET/KD test reported the volunteer's eye movements along the 28 response parameters listed in Table 1. Thus, each test administration for each volunteer resulted in a spreadsheet containing values for these 28 variables. Of these parameters, 10 are averages, 9 are standard deviations, 6 are summed totals, 2 are calculated ratios, and 1 is a rate. It should be noted that the documentation for the ET/KD system did not describe methods, procedures, or algorithms used to define, derive, or calculate these response parameters. Furthermore, the ET/KD did not record the traditional measures of the KDNR test; i.e., the time required to read all 120 numbers or any errors made while reading the numbers. The technician recorded reading time and errors with paper, pencil, and a stopwatch.

Design

Two experiments were conducted. The first experiment compared the two ET/KD systems with 20 volunteers measured twice on each system, alternating between the two systems with half the volunteers starting with one system and the other half starting on the other system in an interleaved fashion. For this experiment, each volunteer's data were collected in a single session. The database of the first experiment consisted of 40 trials on each ET/KD system for a complete dataset of 80 trials. The second experiment evaluated the two ET/KD systems over 5 successive days, with 5 volunteers tested each day twice on both systems, alternating between ET/KD systems to produce a complete dataset of 100 trials.

Table 1. The 28 eye movement output parameters generated by the ET/KD for each test card.

ET/KD Reported Output Variables

Total Fixation Time (seconds)	Standard Deviation Saccade Length (mm)
Average Fixation Time Length (ms)	Total Only Saccade Length In Degrees (deg)
Standard Deviation Fixation Time Length (ms)	Average Saccade Length In Degrees (deg)
Average Fixation Size (mm)	Standard Deviation Saccade Length In Degrees (deg)
Standard Deviation Fixation Size (mm)	Average Saccade Velocity In Degrees (deg/sec)
Average Fixation Polyarea (mm2)	Standard Deviation Saccade Velocity In Degrees
Standard Deviation Fixation Polyarea (mm2)	Total Blinks
Average Inter Fixation Time (ms)	Blink Rate (blinks per minute)
Standard Deviation Inter Fixation Time (ms)	Average Blink Duration (ms)
Saccade Fixation Ratio	Standard Deviation Blink Duration (ms)
Total Saccades (#)	Average Inter Blink Duration (seconds)
Total Saccade Length (mm)	Left Eye Average Pupil Size (mm)
Total Only Saccade Length (mm)	Left Eye Standard Deviation Pupil Size (mm)
Average Saccade Length (mm)	Left Eye Pupillary Stress Ratio

Data Analysis

Descriptive statistics, coefficients of agreements, and differences between individuals were calculated for the 28 ET/KD eye movement output parameters as well as for the pair of conventional KDNR metrics of reading time and number of errors.

Results

Figure 3 displays eye tracking data generated by the ET/KD system for a typical volunteer successively viewing the three test cards. The left half of the figure shows the sequential scan patterns while the right half of the figure shows a heat-type plot in which color identifies the relative length of time the eye's line of sight (LOS) remained in the region of interest. The upper pair of plots displays data obtained with the TC 1. The middle pair displays data obtained with the TC 2. The bottom pair displays data obtained with the TC 3. These data are typical of the results reported for all volunteers. The data presented in the bottom left-right pair of tracings are obviously different from the data presented in the top and middle pair. The eye movements elicited by TC 3 were obviously different from the eye movements associated with either TC 1 or TC 2. In fact, not only are the eye movements different for TC 3, but the characteristic difference between TC 3 and the other two test cards is evident in the TC 3 data tracing: that is, the vertical distance between the rows of fixations or regions of interest is compressed, reflecting the compression of the TC 3. Specifically, the data obtained with TC 3 cannot be confused with the data obtained with either TC 1 or TC 2. On the other hand, there are no obvious differences in the data obtained with TC 1 and TC 2.

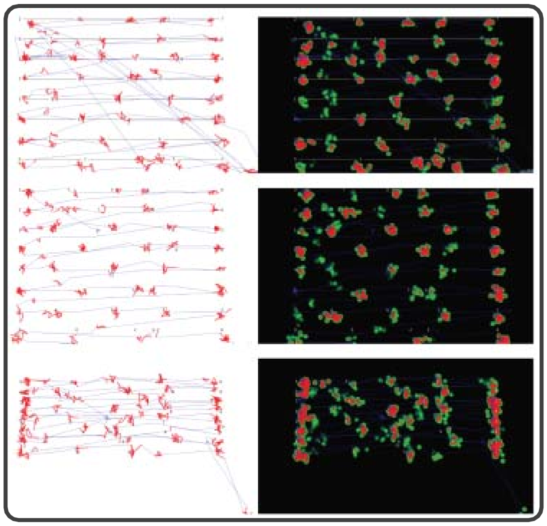


Figure 3. Typical eye tracking plots generated by the ET/KD.

The top left tracing showing the sequential scan patterns for TC 1 displays an unambiguous diagonal slewing of the LOS from the lower right to the upper left. These diagonal eye movements are a result of the test administration procedures. The volunteer starts the test by using the mouse to click a start button that is displayed on the computer screen. This start button is on the lower right corner of the display, so the experiment usually begins with the eye oriented toward the lower right corner, requiring the LOS to slew to the upper left corner to acquire the first digit. The same behavior is evident in the tracking for TC 3. However, the behavior is not as obvious in the data for TC 2, since there is some evidence the diagonal slew of the LOS begins about two thirds of the way to the upper left corner.

For the first experiment, all 20 subjects completed 2 test trials on each ET/KD system. The means for each of the 2 trials for each ET/KD system were calculated and are reported for 21 of the 28 response parameters listed in Table 1. The seven response parameters not reported in Table 2 are means of standard deviations, the interpretation of which some consider problematic.

Table 2. Selected mean measurement values by system and trial number for the first experiment.

20 subjects each completed 2 trials on the two ET/KD systems.

Measure	System 1			System 2		
	Trial 1	**Trial 2**	**Overall**	**Trial 1**	**Trial 2**	**Overall**
Total Time	53.19	50.16	51.68	51.87	49.55	50.71
Total Fixation Time (sec)	13.45	12.45	12.95	13.12	12.42	12.77
Average Fixation Time Length (ms)	226	216.67	221.33	227.25	229.18	228.21
Average Fixation Size (mm)	180.06	172.16	176.11	173.79	176.23	175.01
Average Fixation Polyarea (mm^2)	1156.32	1114.26	1135.29	1090.07	1102.77	1096.42
Average Inter Fixation Time (ms)	286.7	287.75	287.22	289.36	293.22	291.29
Saccade Fixation Ratio	1.5	1.54	1.52	1.54	163.28	1.52
Total Saccades	90.5	88.43	89.47	90.73	84.42	87.58
Total Saccade Length (mm)	32274	31237	31755	31675	29777	30726
Total Only Saccade Length (mm)	21066	20861	20964	21202	19805	20504
Average Saccade Length (mm)	237.48	240.2	238.84	239.3	240.88	240.09
Total Only Saccade Length (deg)	179.12	175.73	177.43	200.1	185.27	192.69
Average Saccade Length (deg)	2.02	2.03	2.02	2.27	2.27	2.27
Total Blinks	6.33	6.05	6.19	7.6	6.77	7.18
Blink Rate (blinks per minute)	21.48	22	21.74	26.44	24.38	25.41
Average Blink Duration (ms)	91.88	70.53	81.21	72.94	85.05	78.99
Average Inter Blink Duration (sec)	2.27	2.12	2.2	1.91	1.78	1.84
Left Eye Average Pupil Size (mm)	2.92	2.93	2.92	2.9	2.85	2.87
Left Eye Pupillary Stress Ratio	22.04	22.46	22.25	23.29	22.03	22.66
Errors	0.3	.035	0.32	0.2	0.3	0.25

Table 3 provides summary statistics, means, and standard deviations (SD) for the 1-day and 5-day experiments. Table 3 has 2 variables more than Table 1 because it includes Total Time and Number of Errors, which were entered by hand into the database. Thus the database contained 30 dependent variables; these were grouped according to 5 general oculometric categories addressing eye blinks, pupillary dynamics, fixations, saccades, and others. Table 3 also includes the dimensions of the response parameters. Blinks are described by 5 response parameters; pupillary dynamics are described by 3 response parameters that address only the left eye; fixations are described by 9 response parameters; saccades are described by 10 response parameters; and lastly, the "other" category contains 3 response parameters, which include the 2 hand-entered variables as well as an undefined variable called the saccade fixation ratio. These summary statistics of the 30 dependent variables were calculated across the 20 or 5 subjects for the 1-day and 5-day experiments, respectively, over TC 1, TC 2, and TC 3, as well as over the two devices and across each administration of the test.

Table 3. Average values for the 1-day and 5-day experiments. The three test cards were averaged for all subjects across all trials over both ET/KD devices.

	Measure	1 day test (Mean ± SD)	5 day test (Mean ± SD)
Blinks	Total Blinks	6.7 ± 6.2	7.8 ± 6.2
	Blink Rate (blinks per minute)	23.8 ± 21.2	31.8 ± 25.2
	Avg Blink Duration (ms)	76.7 ± 97.0	54.9 ± 41.5
	SD Blink Duration (ms)	65.1 ± 92.6	40.0 ± 44.8
	Avg Inter Blink Duration (seconds)	2.0 ± 2.3	1.3 ± 1.1
Pupillary Dynamics	Left Eye Average Pupil Size (mm)	2.9 ± 0.3	2.6 ± 0.3
	Left Eye SD Pupil Size (mm)	0.14 ± 0.05	0.10 ± 0.03
	Left Eye Pupillary Stress Ratio	22.7 ± 7.4	29.9 ± 10.7
Fixations	Total Fixation Time (seconds)	12.4 ± 2.7	9.5 ± 1.7
	Avg Fixation Time Length (ms)	219.1 ± 44.9	163.1 ± 44.1
	SD Fixation Time Length (ms)	143.9 ± 38.8	107.8 ± 22.4
	Avg Fixation Size (inches)	6.8 ± 1.6	5.1 ± 1.3
	SD Fixation Size (mm)	122.9 ± 35.1	99.1 ± 23.9
	Avg Fixation Polyarea (mm2)	1114 ± 315	859 ± 264
	SD Fixation Polyarea (mm2)	1089 ± 357	995 ± 297
	Avg Inter Fixation Time (ms)	283.0 ± 55.2	229.9 ± 47.9
	SD Inter Fixation Time (ms)	169.4 ± 115	141.2 ± 144
Saccades	Total Saccades (#)	87.7 ± 20.5	104.6 ± 39.4
	Total Saccade Length (inches)	1212 ± 212	1177 ± 256
	Total Only Saccade Length (inches)	810.6 ± 124	853.7 ± 192
	Avg Saccade Length (inches)	9.4 ± 1.2	8.5 ± 1.3
	SD Saccade Length (mm)	199.7 ± 42.6	178.1 ± 45.1
	Total Only Saccade Length (deg)	183.2 ± 28.5	183.9 ± 41.9
	Average Saccade Length (deg)	2.1 ± 0.3	1.8 ± 0.3
	SD Saccade Length (deg)	1.8 ± 0.4	1.5 ± 0.4
	Avg Saccade Velocity (deg/sec)	70.4 ± 9.1	61.2 ± 7.6
	SD Saccade Velocity (deg/sec)	55.6 ± 12.5	46.6 ± 9.5
Others	Saccade Fixation Ratio	1.5 ± 0.2	1.7 ± 0.2
	Total Reading Time (seconds)	49.6 ± 8.8	42.7 ± 5.2
	Errors	0.3 ± 0.5	0.0 ± 0.0

Specifically, for the 1-day test, each volunteer was tested twice on each ET/KD system to generate 4 repeated measures. For a volunteer, the TC 1, TC 2, and TC 3 data were averaged for each test administration, and these averages were averaged to generate a single score for each subject. The individual summary data were averaged over the 20 volunteers to generate the means (± SD) displayed in the 1-day testing column of Table 3. A similar strategy was used to collapse the data spanning the 5 days for all 5 volunteers, taking into account the data across three test cards from both systems for each volunteer. These output summary statistics are based on the output of the eye tracking software resident on the ET/KD system.

While the means (± SD) of many of the output response parameters reported in Table 3 seem quite reasonable and have face validity, the values of several response parameters clearly are not. For example, the 1-day and 5-day testing average number of Total Blinks for all three test cards is reported as 6.7 and 7.8, respectively. This is simply not possible with an average reported Blink Rate of 23.8 and 31.8 blinks per minute. Furthermore, the Average Blink Durations (76.7 and 54.9 ms) are extraordinarily brief.

The average pupil size of 2.9 and 2.6 mm is extraordinarily small with almost no evidence of variability (0.14 and 0.10), which may indicate an insensitive rather than a consistent measure. The ‘pupillary stress ratio’ was not defined. Measures of fixations are similarly suspicious. For example, Total Fixation Time (12.4 and 9.5 sec) and Average Fixation Time (219.1 and 163.1 ms) may be reasonable, but the validity of the data is undermined by the fact that the reported average size of the fixations approximate the full size of the display screen (6.8 and 5.1 inches). These are not meaningful measures. Similarly, the reported Average Saccadic Length of between 8 and 9 inches is not meaningful.

Appendices A and B present graphs of the 1-day and 5-day data, respectively, in several different formats to show that the lack of measurement precision is characteristic of these data and not due to a few isolated, idiosyncratic outliers.

Determining Limits of Agreement

An assessment of the agreement between the two ET/KD systems was made for each of the response parameters with the method described by Bland and Altman (Bland & Altman, 1996). In this method, the difference in measurements between the two ET/KD systems (System 2 – System 1) is plotted against the average of the two measurements for every subject. This assessment used the first run on each ET/KD system for each subject. Figure 4 illustrates the procedure with the Total Reading Time response parameter. The middle dotted line is the mean difference in total time for all 20 subjects (-1.3 sec), and the upper and lower dotted lines are the upper and lower limits of agreement (LOAs), respectively. The upper LOA is 6.84 seconds and the lower LOA is -9.49 seconds for total time. The LOAs are calculated by adding and subtracting 1.96 * standard deviation of the differences in measurements, *s*, from the mean difference in measurements, *d*. That is, LOAs = $d \pm 1.96\, s$. Based on this sample, on 95% of the tests, the difference between the two ET/KD systems in Total Reading Time to complete the KDNR test falls between -9.49 and 6.84 seconds. The mean differences and LOAs for each response parameter are given in Table 4. Bland-Altman plots for the other response parameters are given in Appendix A.

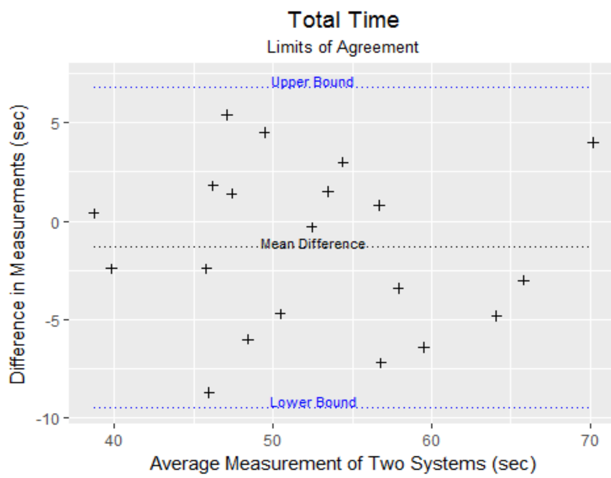


Figure 4. Difference vs mean for Total Reading Time for trial 1 of the first experiment.

Table 4. Measures of agreement for trial 1 of the first experiment.

Measure	**Difference (System 2 – System 1)**	
	Mean	**LOAs**
Total Time (sec)	-1.33	(-9.49, 6.84)
Total Fixation Time (sec)	-0.33	(-3.68, 3.01)
Average Fixation Time Length (ms)	1.25	(-93.70, 96.20)
Average Fixation Size (mm)	-6.27	(-65.43, 52.89)
Average Fixation Polyarea (mm^2)	-66.25	(-541.46, 408.96)
Average Inter Fixation Time (ms)	2.66	(-100.43, 105.75)
Saccade Fixation Ratio	0.04	(-0.35, 0.42)
Total Saccades	0.23	(-45.93, 46.40)
Total Saccade Length (mm)	-598.42	(-9957, 8761)
Total Only Saccade Length (mm)	135.98	(-7213, 7485)
Average Saccade Length (mm)	1.82	(-63.00, 66.63)
Total Only Saccade Length (deg)	20.98	(-38.90, 80.86)
Average Saccade Length (deg)	0.25	(-0.47, 0.96)
Total Blinks	1.27	(-8.15, 10.68)
Blink Rate (blinks/min)	4.95	(-28.28, 38.19)
Average Blink Duration (ms)	-18.94	(-199.60, 161.72)
Average Inter Blink Duration (sec)	-0.37	(-7.54, 6.81)
Left Eye Average Pupil Size (mm)	-0.02	(-0.23, 0.19)
Left Eye Pupillary Stress Ratio	1.25	(-9.60, 12.09)

Calculation of the Coefficient of Agreement

Another method to compare measurements of each response parameter made by the two ET/KD systems is to define a coefficient of agreement (COA), such that a coefficient close to 1.0 signifies good agreement and a coefficient near 0.0 signifies poor agreement. For this purpose, we used the appropriate COA from the options proposed by (Haber & Barnhart, 2008). Since the ET/KD systems were new and untested, we wanted to compare the two systems without considering either system as a reference. Therefore, the COA used is defined as follows:

$$COA = \frac{[MAD(X, X') + MAD(Y, Y')]/2}{MAD(X, Y)}$$

where MAD(X, X') is the mean absolute difference of repeated measurements taken with ET/KD System 1, MAD(Y, Y') is the mean absolute difference of repeated measurements taken with ET/KD System 2, and MAD(X, Y) is the mean absolute difference of the first measurements taken on the two systems. This method makes the numerator the average within-system disagreement and the denominator the between-systems disagreement. The four individual trials for each subject were used to determine this COA.

Using the Total Reading Time measure, MAD(X, X') = 3.565 seconds, MAD(Y, Y') = 3.81 seconds, and MAD(X, Y) = 3.605 seconds. This resulted in a COA of 1.023, which indicates good agreement between the two ET/KD systems. Further evaluation of the agreement between the two systems was made by plotting subject-specific MAD(X, Y) versus [MAD(X, X') + MAD(Y, Y')]/2, which is illustrated in Figure 5. Points near the 45° line indicate good agreement for that subject, while points far away from the line indicate poor subject-specific agreement.

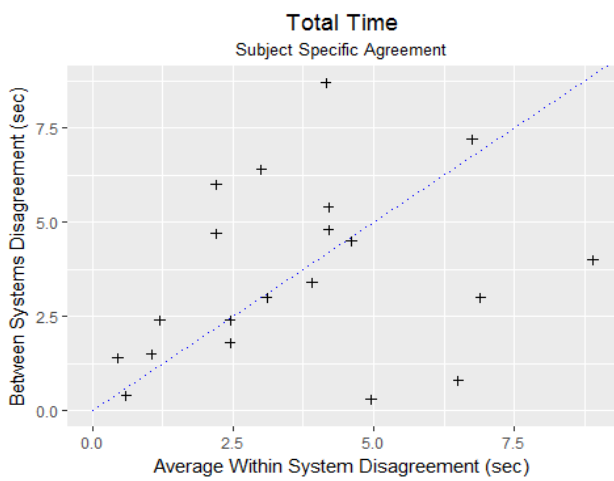


Figure 5. Subject-specific agreement for total response time for the first experiment.

The COAs, as well as the mean, standard deviation, and maximum absolute difference for the response parameters, are shown in Table 5. Subject-specific agreement charts are provided in Appendix B.

Table 5. Measures of agreements of mean differences and COA between the two ET/KD systems for the output parameters for the first experiment.

Measure	Absolute Difference (X, Y)			COA
	mean	sd	max	
Total Time (sec)	3.61	1.53	8.7	1.023
Total Fixation Time (sec)	1.36	1.01	4	1.227
Average Fixation Time Length (ms)	37.51	5.36	128.5	1.018
Average Fixation Size (mm)	24.03	4.26	88.9	1.126
Average Fixation Polyarea (mm^2)	198.86	12.00	622.6	1.078
Average Inter Fixation Time (ms)	38.87	5.79	147.4	1.142
Saccade Fixation Ratio	0.12	0.39	1	1.021
Total Saccades	16.53	4.00	103	1.060
Total Saccade Length (mm)	3483.9	56.12	16488	1.095
Total Only Saccade Length (mm)	2559.9	51.17	13045	1.018
Average Saccade Length (mm)	25.93	4.39	78.6	0.925
Total Only Saccade Length (deg)	27.98	4.84	119.7	0.788
Average Saccade Length (deg)	0.37	0.48	0.9	0.686
Total Blinks	3.57	1.82	14	1.016
Blink Rate (blinks per minute)	12.57	3.44	47.1	1.004
Average Blink Duration (ms)	52.87	8.81	559.9	1.089
Average Inter Blink Duration (sec)	1.81	1.76	17.4	0.990
Left Eye Average Pupil Size (mm)	0.08	0.26	0.3	0.844
Left Eye Pupillary Stress Ratio	4.15	1.91	16.2	1.145

Discussion

The present evaluations of the ET/KD systems were motivated largely because the recent literature widely reports the KDNR test to be a highly sensitive, easy, and reliable sideline sports concussion screener. For example, as of August 2017, a search of Pub Med on the key words “concussion” and “King Devick” produced 52 citations; of which 17 were published during the 8 months of 2017 alone, 11 during 2016, and 10 during 2015. Clearly, during this period there has been a great deal of interest and enthusiasm for the KDNR test as a possible tool for assessing concussion. This literature frequently describes the KDNR test as a measure of eye movements, saccades, or oculomotor behavior, and the KDNR test’s sensitivity to concussion is completely consistent with the well-known sensitivity of eye movements to the disruptive effects of a range of neurological conditions, including traumatic brain injury. The dependence of the KDNR test on eye movements may be logically obvious since a person has to look at the number in order to read it, so the KDNR test seems to offer a simple, intuitive eye movement assessment method using reading behavior measured with paper, pencil, and a stopwatch. The KDNR test is eminently practical and easy to use as a possible screener in a range of sports settings from football and rugby to mixed martial arts (MMA) and boxing, just to name a few contexts in which the KDNR test has been used with extraordinarily good results. A recent report describes its use in a telemedicine context (Vargas et al., 2017). The KDNR test is widely reported to have high sensitivity and specificity for sports-related concussion when the test is used properly; that is, the metric is a difference score from an individually determined baseline. Therefore, the KDNR test ideally needs to be administered before and after the suspected concussive event.

This stipulation may be a complicating factor limiting KDNR use with unpredictable concussive events.

Despite the fact that the KDNR test is widely described as a measure of eye movements or oculomotor behavior, the literature reporting objectively measured eye movements during the KDNR test is extremely thin, consisting at the time of this writing of only four published papers to our knowledge. In two of these papers the measurements were made in volunteers with some neurological condition such as concussion (Rizzo, Hudson, Dai, Birkemeier, et al., 2016) and multiple sclerosis (Hainline et al., 2017), and a third paper reports data from normal control volunteers (Rizzo, Hudson, Dai, Desai, et al., 2016). These three papers are from the same laboratory using the same objective infra-red eye tracking technology. Another research group recorded eye movements during KDNR test performance in normal volunteers while they were exposed to normobaric hypoxia (Stepanek et al., 2013). This study was intended to assess the neurocognitive effects of hypoxia in order to identify possible leading indicators of hypoxia-induced neurocognitive deficits. Thus, it seems that there is very little actual empirical information describing the relationship between eye movements and KDNR test performance. On one hand, this lack of information may be relatively moot for the use of the KDNR test as a practical concussion screener in the field where sensitivity may be more important than specificity. On the other hand, this situation means that a change in the KDNR test performance is widely attributed to changes in oculomotor behavior without actually quantifying or describing the oculomotor behavior and associated changes caused by concussion.

For example, the details of eye movements made while reading can be quantified in a great number of ways. The ET/KD system assessed here listed more than 28; disruption of almost any of these could contribute to a change in KDNR test performance. This level of detail about eye movements is invisible to the KDNR test; unfortunately, this level of detail about eye movements seems to be equally invisible to much of the discussions about the KDNR test, since it is so widely described simply as a test of eye movements. The KDNR test depends on a complex chain of behavior that culminates in an explicit verbalization. Performance is as dependent on oculomotor behavior and retinal and central visual sensory and perceptional processes as it is on motivation, cognition, and attention, as well as speech control and cardiopulmonary physiology.

A recent incidental observation made while integrating the KDNR test with our cardiopulmonary and vascular measures made us sensitive to the assumption that the KDNR test is a measure of eye movements (Temme et al., 2017). During normobaric hypoxia, we found obvious and dramatic alterations in breathing behavior that had nothing to do with eye movements as such and all to do with breath control. These were dramatic and obvious effects, and in retrospect should not have been surprising. The availability of oxygen directly affected the individual's speaking behavior. This incidental observation suggested the possibility that factors which are typically ignored completely can significantly alter KDNR test performance. This possibility further implies that the KDNR test as a measure of the effects of concussion on eye movements rests on the usually tacit assumption that all other components contributing to reading other than oculomotor behavior are constant. In other words, concussion affects nothing but the oculomotor system and its behavior. It is possible, for example, that the normal patterns of speech production are altered by concussion. Hence, it is important to measure eye movements objectively as well as other visual and behavioral components of reading.

The initial intention of these studies was to evaluate the reliability, repeatability, and comparability of the data obtained with the two ET/KD systems. The statistical approach was to use the Bland-Altman graphical limits of agreement method as well as the coefficient of agreement statistics developed by Haber and Barnhart (2008). However, because of the obvious lack of face validity of several of the reported measures (Tables 2 and 3), it seemed pointless to further pursue the questions of reliability, repeatability, and comparability, particularly since the evaluation of reliability, etc., depends on a precision of measurement that these two ET/KD systems do not support.

Eye tracking remains a highly complicated, technologically demanding area of research, which is one of the reasons the KDNR test is so popular; it is easy to use in a chaotic field environment, quick to administer, and simple to score. For these reasons, we were interested in implementing the COTS ET/KD system when it became available. We expected the ET/KD system to be a valid and reliable turnkey system which, unfortunately, turned out not to be the case, as the results of the present evaluations demonstrate. The two ET/KD systems are not usable in that they did not provide meaningful information. These findings are extraordinarily important because they prevented these ET/KD systems from being deployed for research purposes. Their use in research could have produced misleading results, which could have been an immense waste of time and money. When we discussed these results with the manufacturer, the manufacturer did not demur but readily agreed, blaming the system's shortcoming on the third-party commercial eye tracker used in fabricating the system. That is, the manufacturer of the ET/KD systems we purchased acknowledged that the ET/KD systems were inadequate, while also asserting that the third-party eye tracker used in assembling the ET/KD system was the source of the shortcoming. The manufacturer insouciantly added that a better eye tracker is now an upgrade we can purchase, and the new proprietary software version would be provided to us for free if we wished; an offer we declined. We are currently using a different eye tracking technology to support our research.

The most important conclusion of this work is that systematic test and evaluation of new equipment is essential before the new equipment can be trusted for research purposes. This is particularly important when there is limited information, or little to no publication record associated with the equipment, or available information predominantly comes from sources with a conflicting financial interest. Unfortunately, the present study, as necessary as it was, consumed substantial resources and did not appreciably advance science; but at least the study identified a potential impediment and misinformation. We recommend that the manufacturer should report formal evaluation before the device is marketed as a research tool, and failures to do this should also be documented.

References

Balaban, C., Hoffer, M. E., Szczupak, M., Snapp, H., Crawford, J., Murphy, S., . . . Kiderman, A. (2016). Oculomotor, Vestibular, and Reaction Time Tests in Mild Traumatic Brain Injury. *PLoS One, 11*(9), e0162168. doi:10.1371/journal.pone.0162168

Barker, T., Russo, S. A., Barker, G., Rice, M. A., Jr., Jeffrey, M. G., Broderick, G., & Craddock, T. J. A. (2017). A case matched study examining the reliability of using ImPACT to assess effects of multiple concussions. *BMC Psychol, 5*(1), 14. doi:10.1186/s40359-017-0184-1

Barnes, G. R. (2011). Ocular pursuit movements. In S. P. Liversedge, I. D. Gilchrest, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press.

Bland, J. M., & Altman, D. G. (1996). Statistics notes: measurement error proportional to the mean. *Bmj, 313*(7049), 106.

Boyer, E. O., Portron, A., Bevilacqua, F., & Lorenceau, J. (2017). Continuous Auditory Feedback of Eye Movements: An Exploratory Study toward Improving Oculomotor Control. *Front Neurosci, 11*, 197. doi:10.3389/fnins.2017.00197

Brahm, K. D., Wilgenburg, H. M., Kirby, J., Ingalla, S., Chang, C. Y., & Goodrich, G. L. (2009). Visual impairment and dysfunction in combat-injured servicemembers with traumatic brain injury. *Optom Vis Sci, 86*(7), 817-825. doi:10.1097/OPX.0b013e3181adff2d

Broglio, S. P., McCrea, M., McAllister, T., Harezlak, J., Katz, B., Hack, D., . . . Investigators, C. C. (2017). A National Study on the Effects of Concussion in Collegiate Athletes and US Military Service Academy Members: The NCAA-DoD Concussion Assessment, Research and Education (CARE) Consortium Structure and Methods. *Sports Med*. doi:10.1007/s40279-017-0707-1

Capo-Aponte, J. E., Jorgensen-Wagers, K. L., Sosa, J. A., Walsh, D. V., Goodrich, G. L., Temme, L. A., & Riggs, D. W. (2016). Visual Dysfunctions at Different Stages After Blast and Non-blast Mild Traumatic Brain Injury. *Optom Vis Sci*. doi:10.1097/OPX.0000000000000825

Cifu, D. X., Wares, J. R., Hoke, K. W., Wetzel, P. A., Gitchel, G., & Carne, W. (2015). Differential eye movements in mild traumatic brain injury versus normal controls. *J Head Trauma Rehabil, 30*(1), 21-28. doi:10.1097/HTR.0000000000000036

Davenport, N. D. (2016). The Chaos of Combat: An Overview of Challenges in Military Mild Traumatic Brain Injury Research. *Front Psychiatry, 7*, 85. doi:10.3389/fpsyt.2016.00085

Davies, E. C., Henderson, S., Balcer, L. J., & Galetta, S. L. (2012). Residency training: the King-Devick test and sleep deprivation: study in pre- and post-call neurology residents. *Neurology, 78*(17), e103-106. doi:10.1212/WNL.0b013e318251833d

Dziemianowicz, M. S., Kirschen, M. P., Pukenas, B. A., Laudano, E., Balcer, L. J., & Galetta, S. L. (2012). Sports-related concussion testing. *Curr Neurol Neurosci Rep, 12*(5), 547-559. doi:10.1007/s11910-012-0299-y

Galetta, K. M., Barrett, J., Allen, M., Madda, F., Delicata, D., Tennant, A. T., . . . Balcer, L. J. (2011). The King-Devick test as a determinant of head trauma and concussion in boxers and MMA fighters. *Neurology, 76*(17), 1456-1462. doi:10.1212/WNL.0b013e31821184c9

Galetta, K. M., Brandes, L. E., Maki, K., Dziemianowicz, M. S., Laudano, E., Allen, M., . . . Balcer, L. J. (2011). The King-Devick test and sports-related concussion: study of a rapid visual screening tool in a collegiate cohort. *J Neurol Sci, 309*(1-2), 34-39. doi:10.1016/j.jns.2011.07.039

Galetta, K. M., Morganroth, J., Moehringer, N., Mueller, B., Hasanaj, L., Webb, N., . . . Balcer, L. J. (2015). Adding Vision to Concussion Testing: A Prospective Study of Sideline Testing in Youth and Collegiate Athletes. *J Neuroophthalmol*. doi:10.1097/WNO.0000000000000226

Galetta, M. S., Galetta, K. M., McCrossin, J., Wilson, J. A., Moster, S., Galetta, S. L., . . . Master, C. L. (2013). Saccades and memory: baseline associations of the King-Devick and SCAT2 SAC tests in professional ice hockey players. *J Neurol Sci, 328*(1-2), 28-31. doi:10.1016/j.jns.2013.02.008

Gilchrist, I. (2011). Saccades. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford Handboook of Eye Movements*. Oxford: Oxford University Press.

Haber, M., & Barnhart, H. X. (2008). A general approach to evaluating agreement between two observers or methods of measurement from quantitative data with replicated measurements. *Statistical Methods in Medical Research, 17*(2), 151-169.

Hainline, C., Rizzo, J. R., Hudson, T. E., Dai, W., Birkemeier, J., Raynowska, J., . . . Rucker, J. C. (2017). Capturing saccades in multiple sclerosis with a digitized test of rapid number naming. *J Neurol, 264*(5), 989-998. doi:10.1007/s00415-017-8484-1

Handmaker, H., & Waldorf, R. A. (2013). Comment: the King-Devick test and sports-related concussion: study of a rapid visual screening tool in a collegiate cohort. *J Neurol Sci, 327*(1-2), 80. doi:10.1016/j.jns.2013.01.023

Heitger, M. H., Jones, R. D., Macleod, A. D., Snell, D. L., Frampton, C. M., & Anderson, T. J. (2009). Impaired eye movements in post-concussion syndrome indicate suboptimal brain function beyond the influence of depression, malingering or intellectual ability. *Brain, 132*(Pt 10), 2850-2870. doi:10.1093/brain/awp181

Hoge, C. W., McGurk, D., Thomas, J. L., Cox, A. L., Engel, C. C., & Castro, C. A. (2008). Mild traumatic brain injury in U.S. Soldiers returning from Iraq. *N Engl J Med, 358*(5), 453-463. doi:10.1056/NEJMoa072972

Iverson, G. L., Echemendia, R. J., Lamarre, A. K., Brooks, B. L., & Gaetz, M. B. (2012). Possible lingering effects of multiple past concussions. *Rehabil Res Pract, 2012*, 316575. doi:10.1155/2012/316575

Kennedy, J. E., Jaffee, M. S., Leskin, G. A., Stokes, J. W., Leal, F. O., & Fitzpatrick, P. J. (2007). Posttraumatic stress disorder and posttraumatic stress disorder-like symptoms and mild traumatic brain injury. *J Rehabil Res Dev, 44*(7), 895-920.

Khodaee, M., Currie, D. W., Asif, I. M., & Comstock, R. D. (2016). Nine-year study of US high school soccer injuries: data from a national sports injury surveillance programme. *Br J Sports Med*. doi:10.1136/bjsports-2015-095946

Leigh, R. J., & Zee, D. S. (2006). *The Neurology of Eye Movements*. Oxford Oxford University Press.

Leong, D. F., Balcer, L. J., Galetta, S. L., Liu, Z., & Master, C. L. (2014). The King-Devick test as a concussion screening tool administered by sports parents. *J Sports Med Phys Fitness, 54*(1), 70-77.

Lin, T. P., Adler, C. H., Hentz, J. G., Balcer, L. J., Galetta, S. L., & Devick, S. (2014). Slowing of number naming speed by King-Devick test in Parkinson's disease. *Parkinsonism Relat Disord, 20*(2), 226-229. doi:10.1016/j.parkreldis.2013.10.009

Lin, T. P., Rigby, H., Adler, J. S., Hentz, J. G., Balcer, L. J., Galetta, S. L., . . . Adler, C. H. (2014). Abnormal Visual Contrast Acuity in Parkinson's Disease. *J Parkinsons Dis*. doi:10.3233/JPD-140470

Liversedge, S., Gilchrist, I., & Everling, S. (2011). *The Oxford handbook of eye movements*: Oxford University Press.

Maguen, S., Lau, K. M., Madden, E., & Seal, K. (2012). Relationship of screen-based symptoms for mild traumatic brain injury and mental health problems in Iraq and Afghanistan veterans: Distinct or overlapping symptoms? *J Rehabil Res Dev, 49*(7), 1115-1126.

Marion, D. W., Curley, K. C., Schwab, K., & Hicks, R. R. (2011). Proceedings of the military mTBI Diagnostics Workshop, St. Pete Beach, August 2010. *J Neurotrauma, 28*(4), 517-526. doi:10.1089/neu.2010.1638

Marshall, S. W., Guskiewicz, K. M., Shankar, V., McCrea, M., & Cantu, R. C. (2015). Epidemiology of sports-related concussion in seven US high school and collegiate sports. *Inj Epidemiol, 2*(1), 13. doi:10.1186/s40621-015-0045-4

Martinez-Conde, S., & Macknik, S. L. (2011). Microsaccades. In S. P. Liversedge, I. D. Gilchrest, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press.

Molloy, J. H., Murphy, I., & Gissane, C. (2017). The King-Devick (K-D) test and concussion diagnosis in semi-professional rugby union players. *J Sci Med Sport*. doi:10.1016/j.jsams.2017.02.002

Moster, S., Wilson, J. A., Galetta, S. L., & Balcer, L. J. (2014). The King-Devick (K-D) test of rapid eye movements: a bedside correlate of disability and quality of life in MS. *J Neurol Sci, 343*(1-2), 105-109. doi:10.1016/j.jns.2014.05.047

Munce, T. A., Dorman, J. C., Odney, T. O., Thompson, P. A., Valentine, V. D., & Bergeron, M. F. (2014). Effects of youth football on selected clinical measures of neurologic function: a pilot study. *J Child Neurol, 29*(12), 1601-1607. doi:10.1177/0883073813509887

Padula, W. V., Capo-Aponte, J. E., Padula, W. V., Singman, E. L., & Jenness, J. (2017). The consequence of spatial visual processing dysfunction caused by traumatic brain injury (TBI). *Brain Inj*, 1-12. doi:10.1080/02699052.2017.1291991

Regasa, L. E., Thomas, D. M., Gill, R. S., Marion, D. W., & Ivins, B. J. (2016). Military Deployment May Increase the Risk for Traumatic Brain Injury Following Deployment. *J Head Trauma Rehabil, 31*(1), E28-35. doi:10.1097/HTR.0000000000000155

Rizzo, J. R., Hudson, T. E., Dai, W., Birkemeier, J., Pasculli, R. M., Selesnick, I., . . . Rucker, J. C. (2016). Rapid number naming in chronic concussion: eye movements in the King-Devick test. *Ann Clin Transl Neurol, 3*(10), 801-811. doi:10.1002/acn3.345

Rizzo, J. R., Hudson, T. E., Dai, W., Desai, N., Yousefi, A., Palsana, D., . . . Rucker, J. C. (2016). Objectifying eye movements during rapid number naming: Methodology for assessment of normative data for the King-Devick test. *J Neurol Sci, 362*, 232-239. doi:10.1016/j.jns.2016.01.045

Schmid, K. E., & Tortella, F. C. (2012). The diagnosis of traumatic brain injury on the battlefield. *Front Neurol, 3*, 90. doi:10.3389/fneur.2012.00090

Seal, K. H., Bertenthal, D., Samuelson, K., Maguen, S., Kumar, S., & Vasterling, J. J. (2016). Association between mild traumatic brain injury and mental health problems and self-reported cognitive dysfunction in Iraq and Afghanistan Veterans. *J Rehabil Res Dev, 53*(2), 185-198. doi:10.1682/JRRD.2014.12.0301

Stepanek, J., Cocco, D., Pradhan, G. N., Smith, B. E., Bartlett, J., Studer, M., . . . Cevette, M. J. (2013). Early detection of hypoxia-lnduced cognitive impairment using the King-Devick test. *Aviat Space Environ Med, 84*(10), 1017-1022.

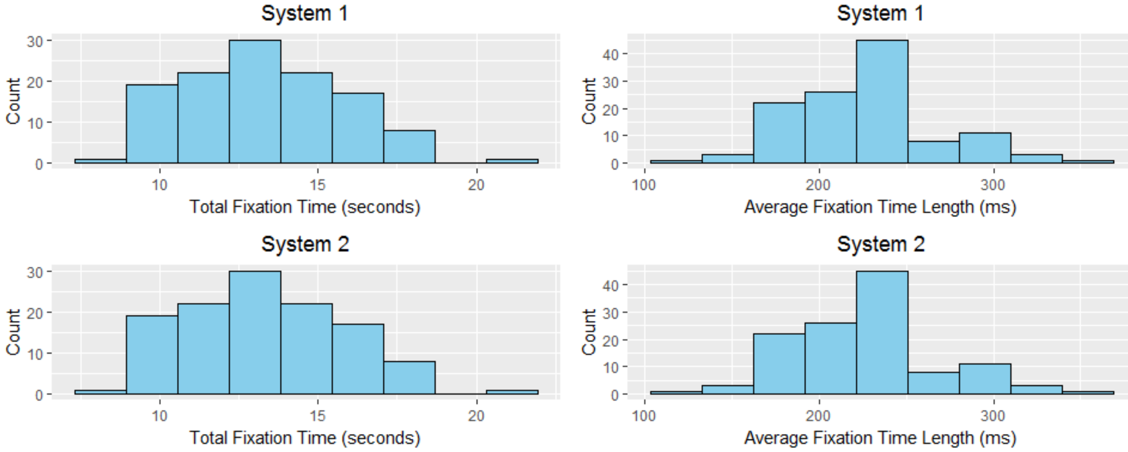
Temme, L. A., St Onge, P., Adams, M., Still, D. L., Statz, J. K., & Williams, S. T. (2017). A Novel, Inexpensive Method to Monitor, Record, and Analyze Breathing Behavior During Normobaric Hypoxia Generated by the Reduced Oxygen Breathing Device. *Mil Med, 182*(S1), 210-215. doi:10.7205/MILMED-D-16-00053

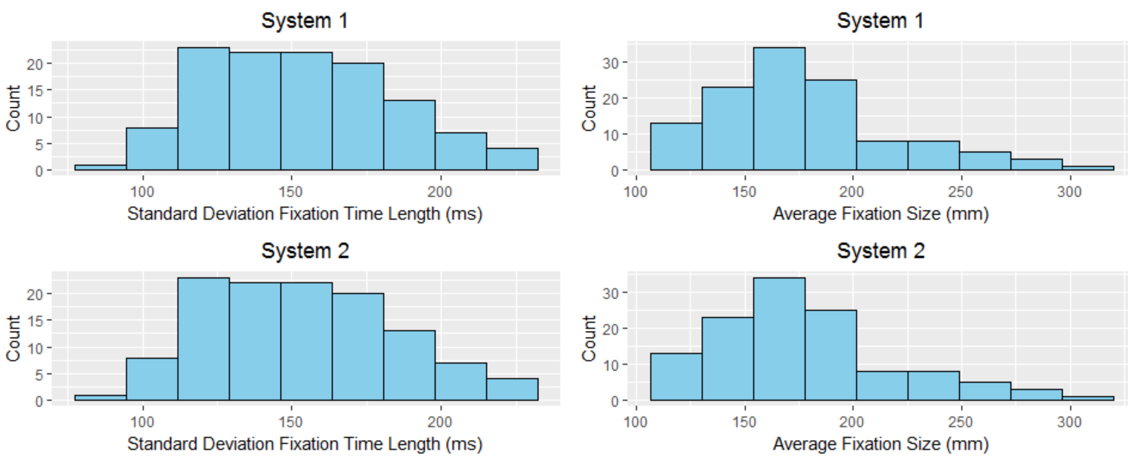
Vargas, B. B., Shepard, M., Hentz, J. G., Kutyreff, C., Hershey, L. G., & Starling, A. J. (2017). Feasibility and accuracy of teleconcussion for acute evaluation of suspected concussion. *Neurology, 88*(16), 1580-1583.

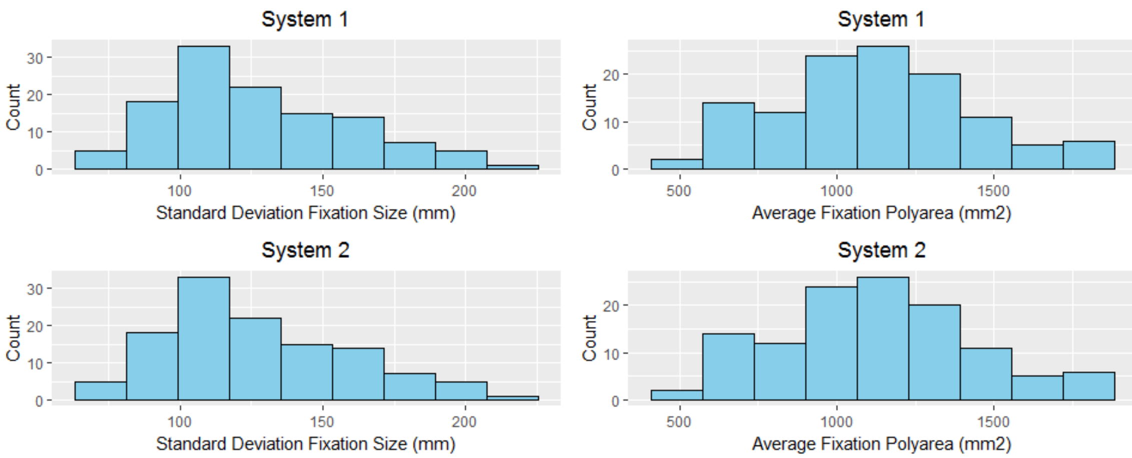
Ventura, R. E., Balcer, L. J., Galetta, S. L., & Rucker, J. C. (2016). Ocular motor assessment in concussion: Current status and future directions. *J Neurol Sci, 361*, 79-86. doi:10.1016/j.jns.2015.12.010

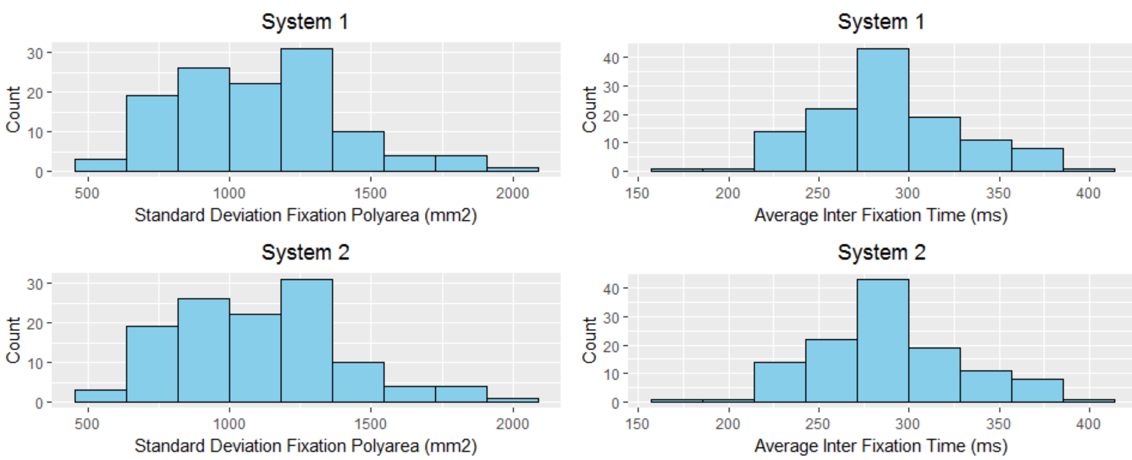
This page is intentionally blank.

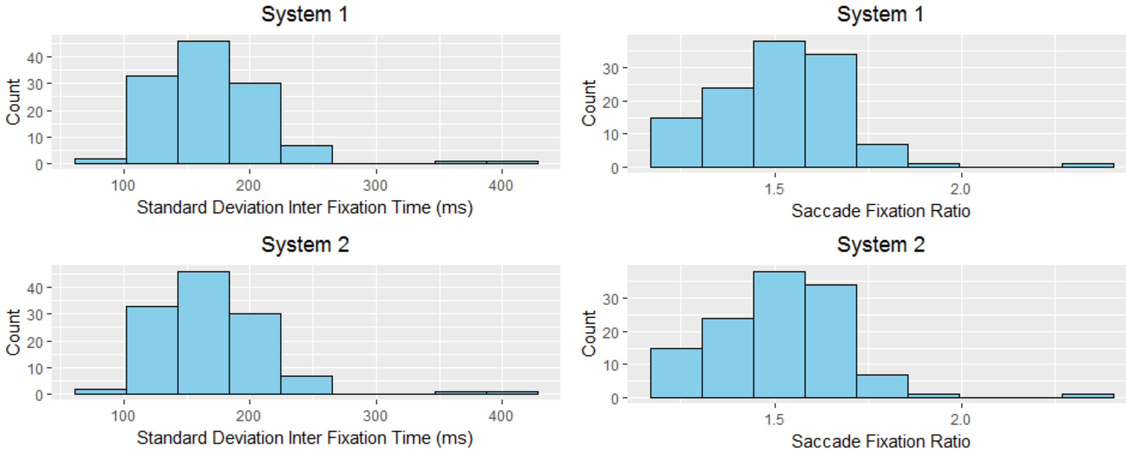
Appendix A. Bland-Altman plots for the response parameters

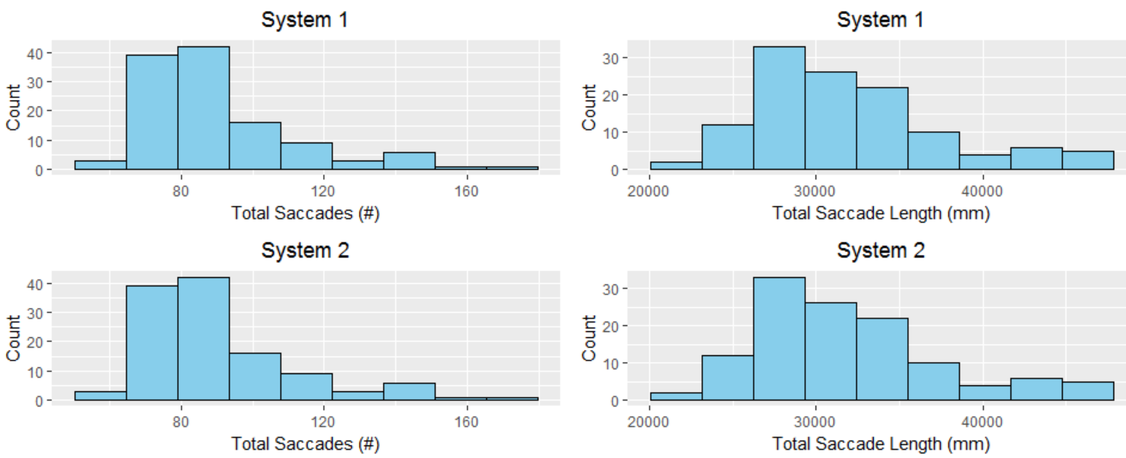


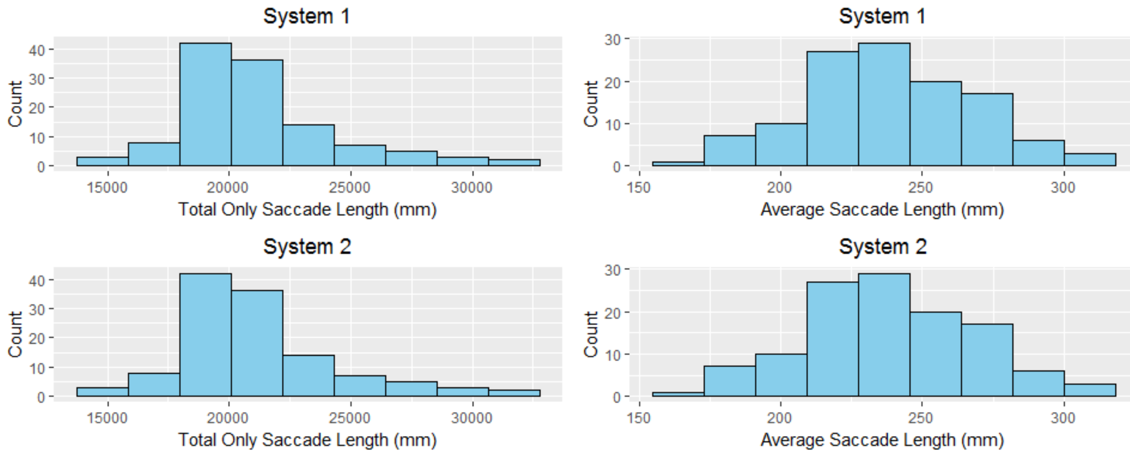


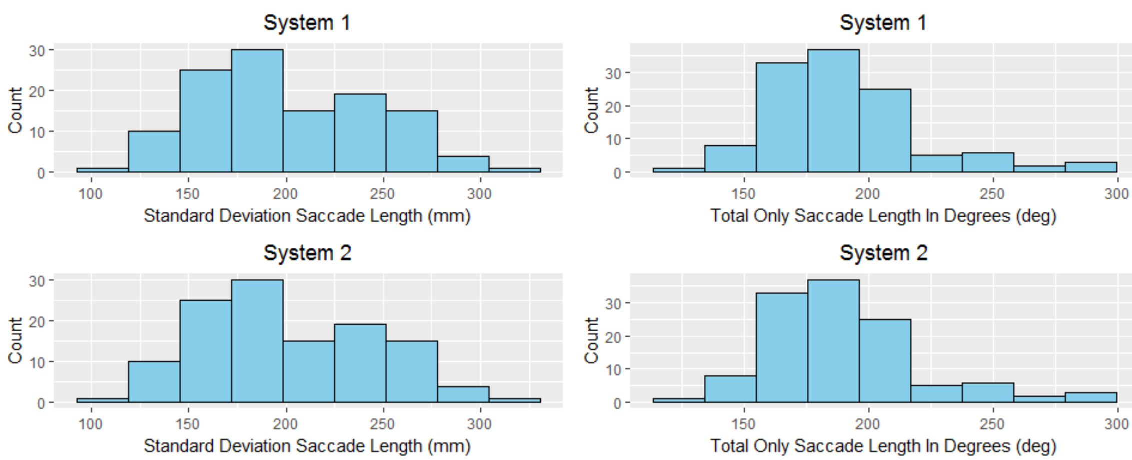

System 1
Count
Standard Deviation Fixation Size (mm)
System 1
Count
Average Fixation Polyarea (mm2)
System 2
Count
Standard Deviation Fixation Size (mm)
System 2
Count
Average Fixation Polyarea (mm2)

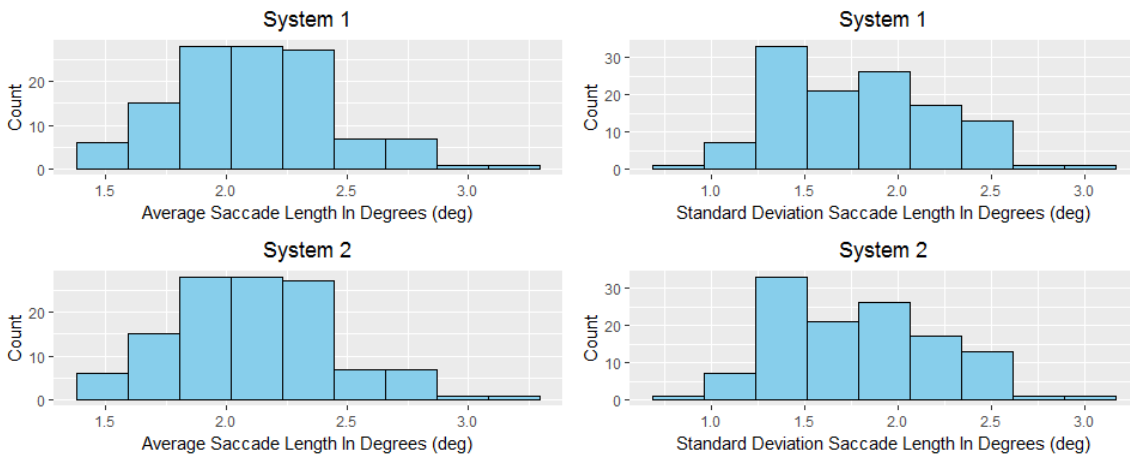

System 1
Count
Standard Deviation Fixation Polyarea (mm2)
System 1
Count
Average Inter Fixation Time (ms)
System 2
Count
Standard Deviation Fixation Polyarea (mm2)
System 2
Count
Average Inter Fixation Time (ms)

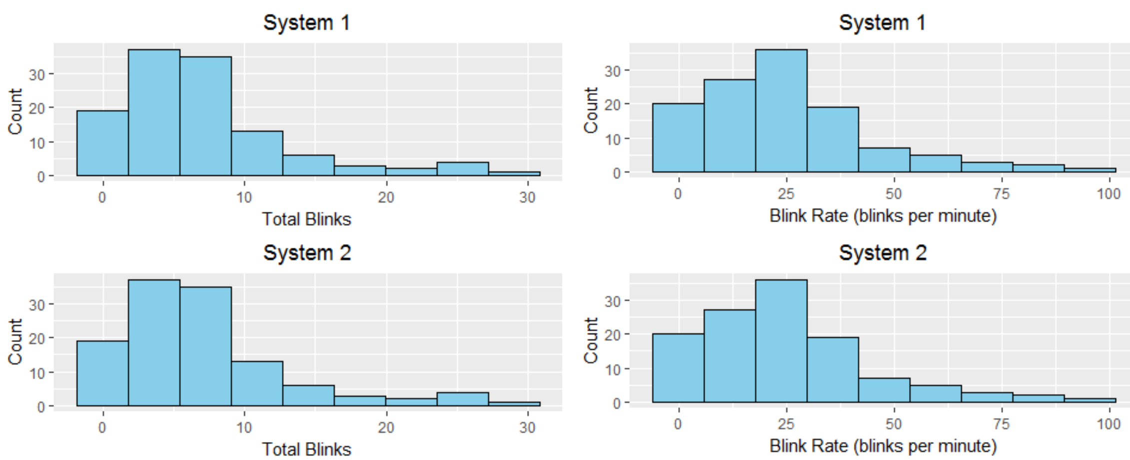

System 1
Count
Standard Deviation Inter Fixation Time (ms)
System 2
Count
Standard Deviation Inter Fixation Time (ms)
System 1
Count
Saccade Fixation Ratio
System 2
Count
Saccade Fixation Ratio

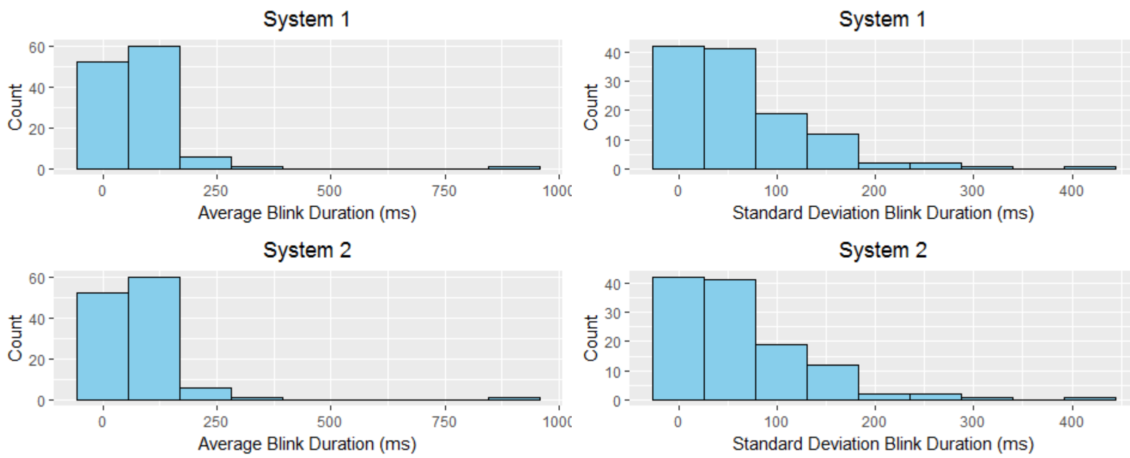

System 1
Count
Total Saccades (#)
System 2
Count
Total Saccades (#)
System 1
Count
Total Saccade Length (mm)
System 2
Count
Total Saccade Length (mm)

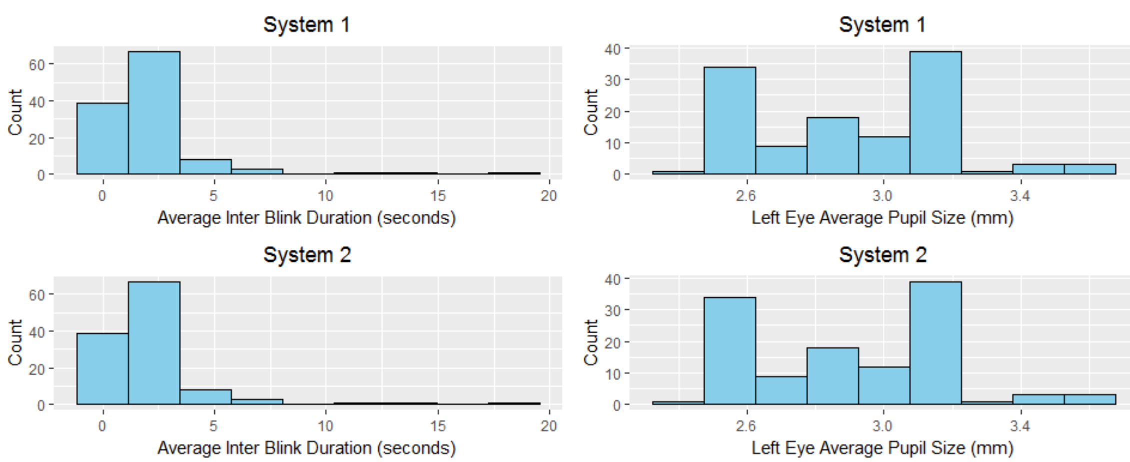

System 1
Count
Total Only Saccade Length (mm)
System 1
Count
Average Saccade Length (mm)
System 2
Count
Total Only Saccade Length (mm)
System 2
Count
Average Saccade Length (mm)

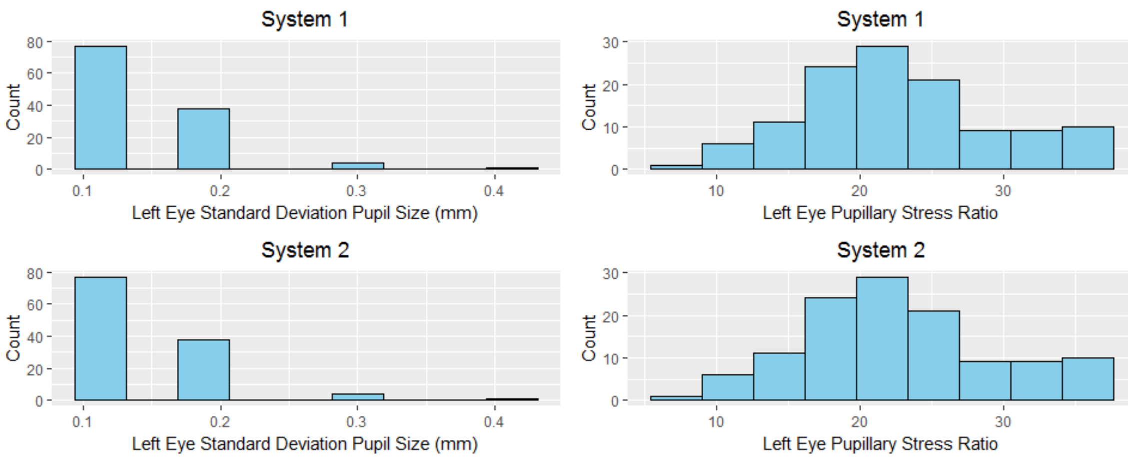

System 1
Count
Standard Deviation Saccade Length (mm)
System 1
Count
Total Only Saccade Length In Degrees (deg)
System 2
Count
Standard Deviation Saccade Length (mm)
System 2
Count
Total Only Saccade Length In Degrees (deg)

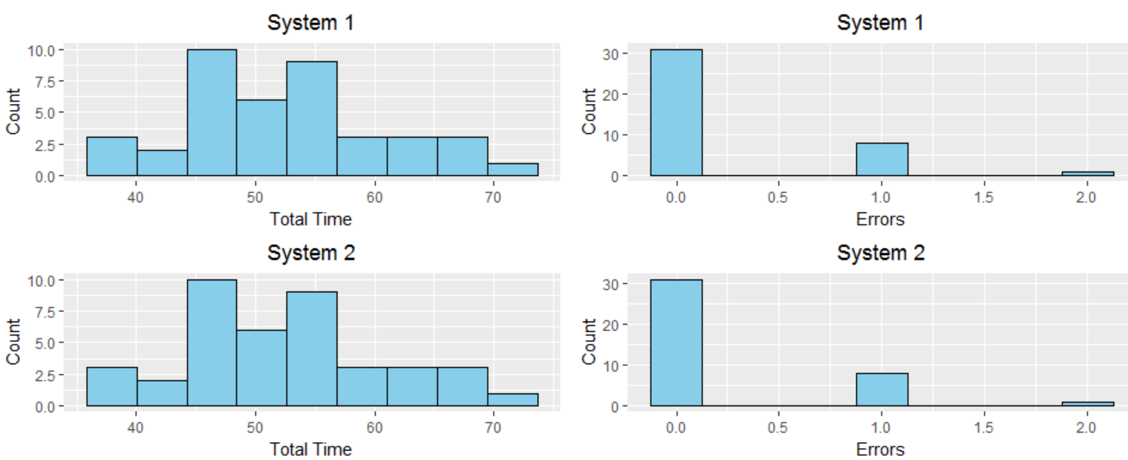

System 1
Count
Average Saccade Length In Degrees (deg)
System 1
Count
Standard Deviation Saccade Length In Degrees (deg)
System 2
Count
Average Saccade Length In Degrees (deg)
System 2
Count
Standard Deviation Saccade Length In Degrees (deg)


System 1
Count
Total Blinks
System 1
Count
Blink Rate (blinks per minute)
System 2
Count
Total Blinks
System 2
Count
Blink Rate (blinks per minute)

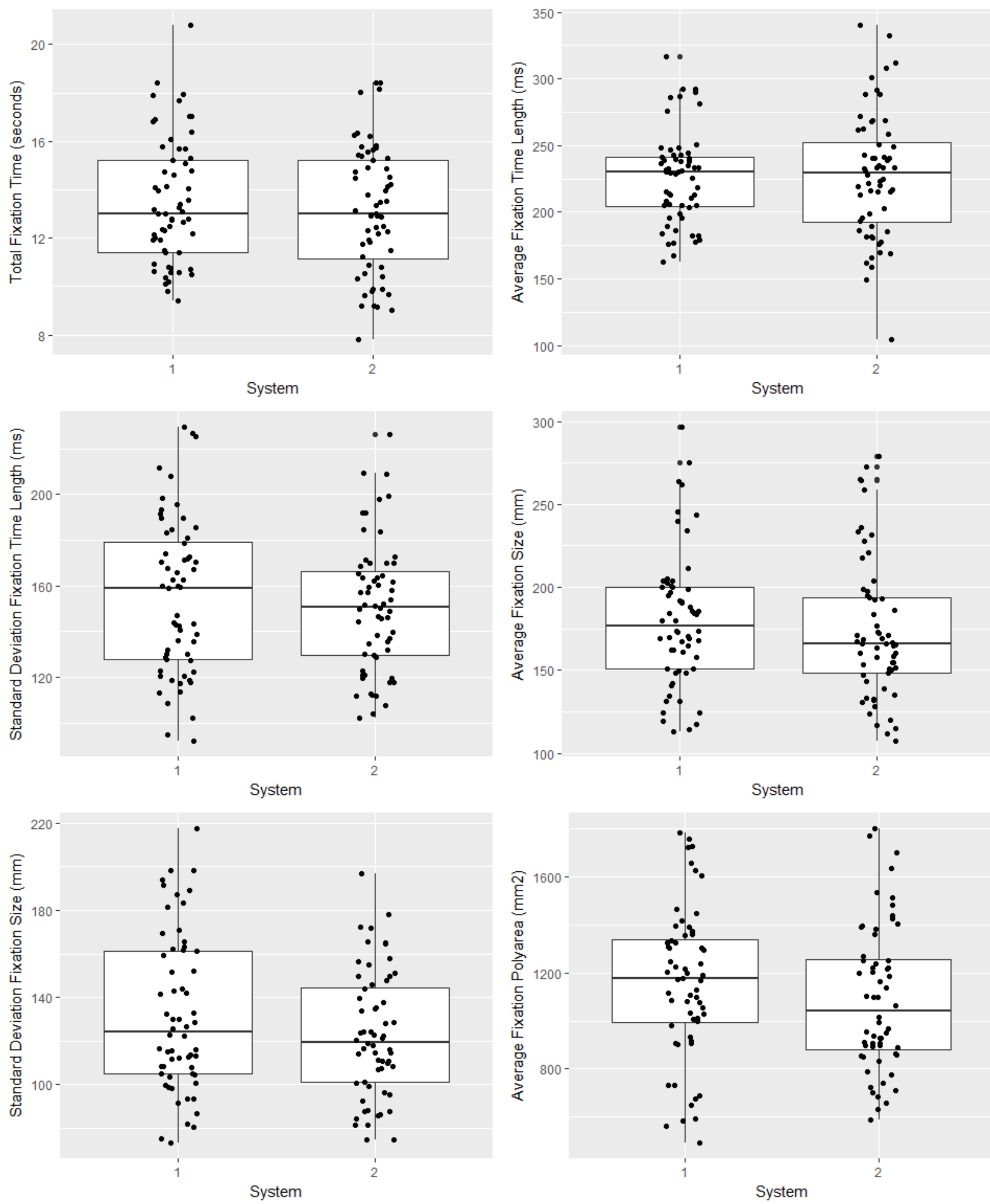

System 1
Count
Average Blink Duration (ms)
System 1
Count
Standard Deviation Blink Duration (ms)
System 2
Count
Average Blink Duration (ms)
System 2
Count
Standard Deviation Blink Duration (ms)

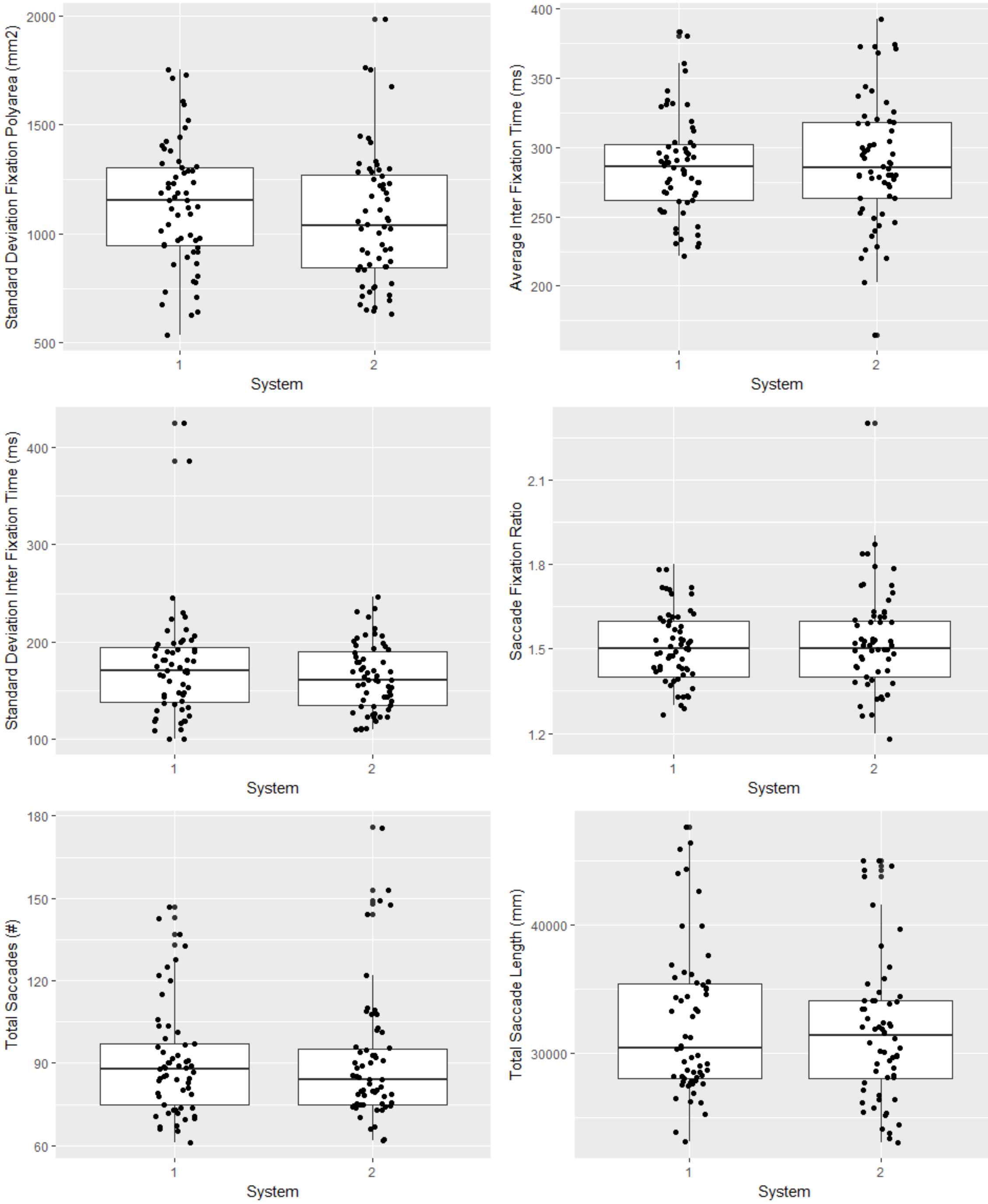

System 1
Count
Average Inter Blink Duration (seconds)
System 1
Count
Left Eye Average Pupil Size (mm)
System 2
Count
Average Inter Blink Duration (seconds)
System 2
Count
Left Eye Average Pupil Size (mm)

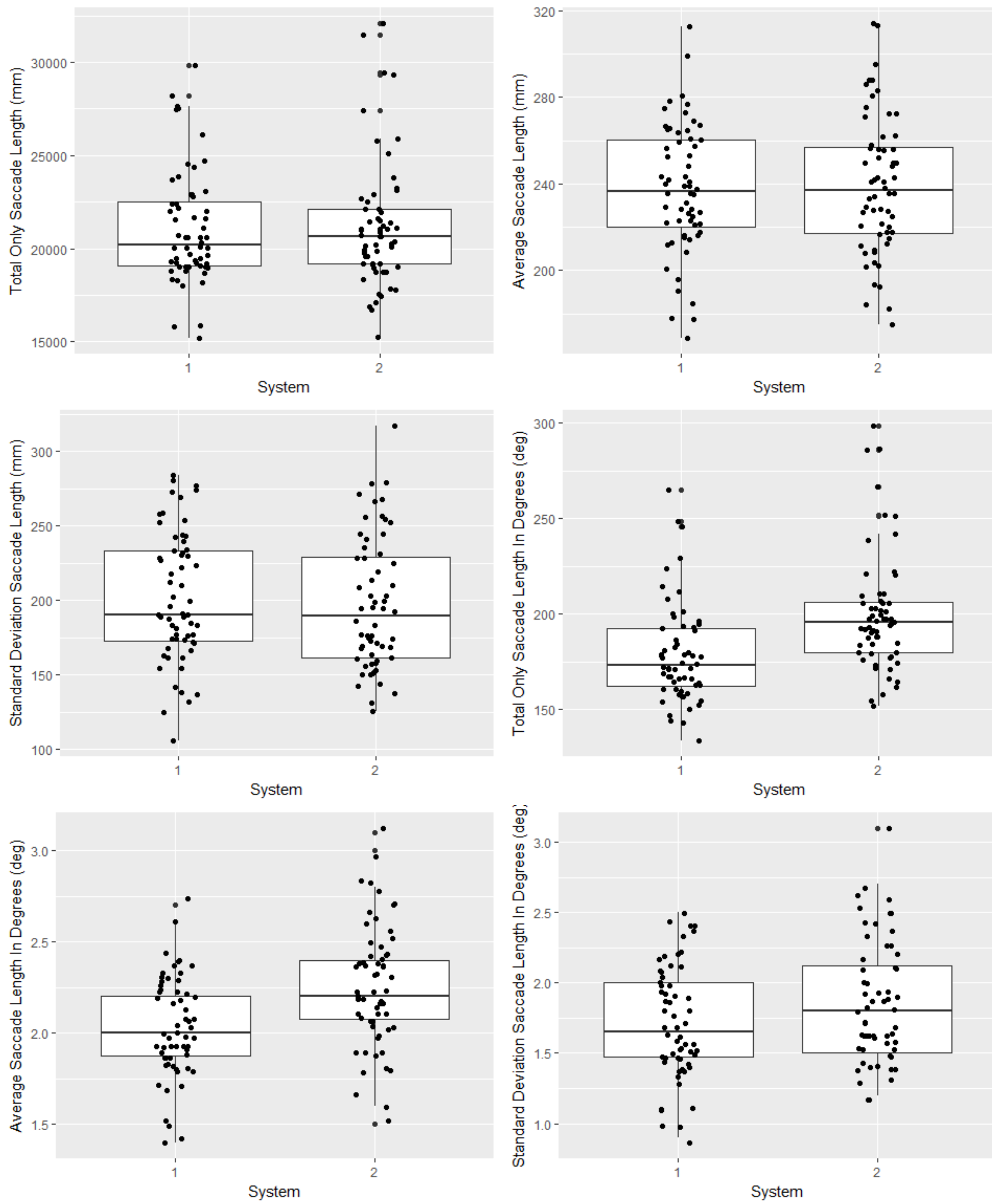

System 1
Count
Left Eye Standard Deviation Pupil Size (mm)
System 1
Count
Left Eye Pupillary Stress Ratio
System 2
Count
Left Eye Standard Deviation Pupil Size (mm)
System 2
Count
Left Eye Pupillary Stress Ratio

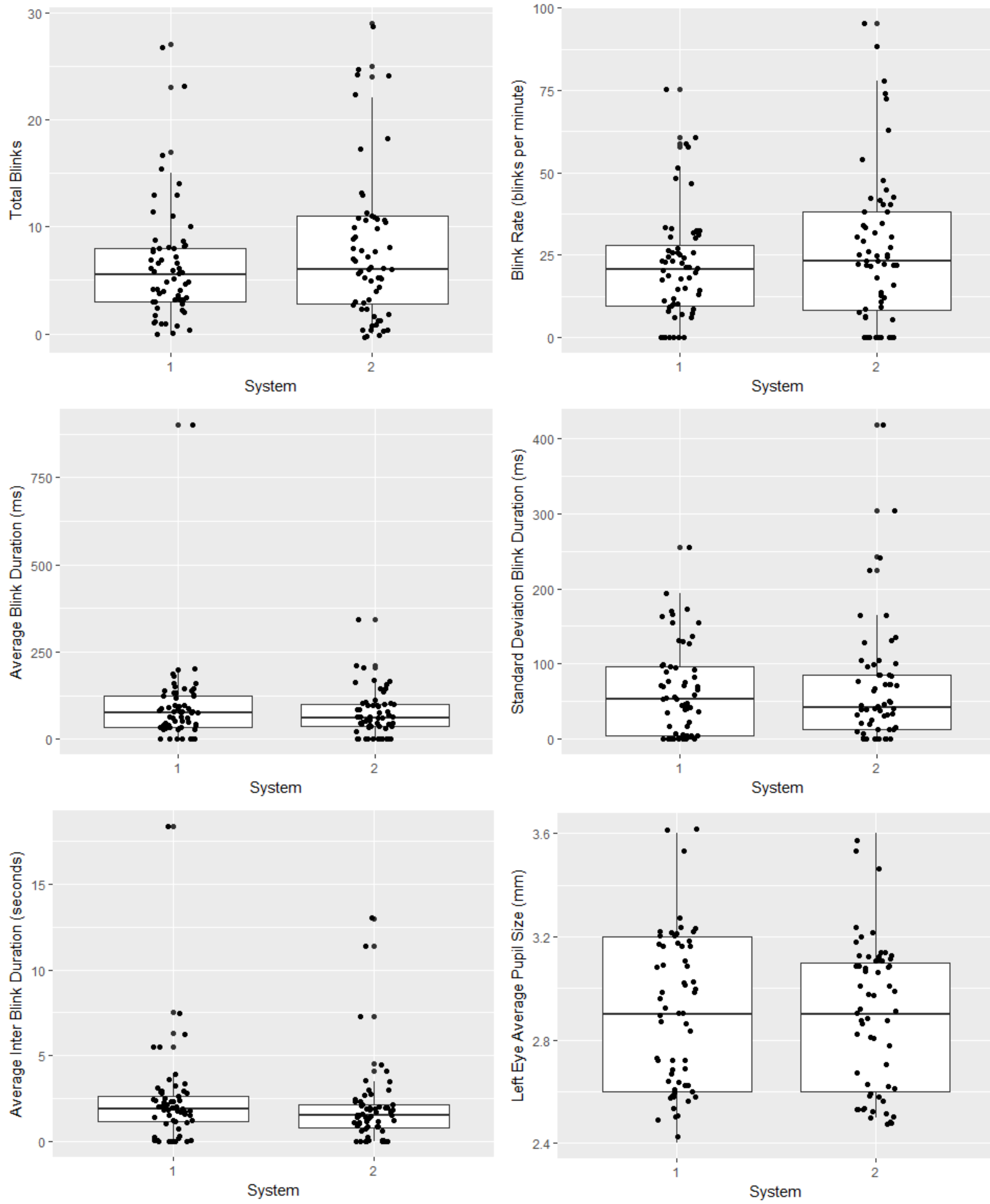

System 1
Count
Total Time
System 1
Count
Errors
System 2
Count
Total Time
System 2
Count
Errors

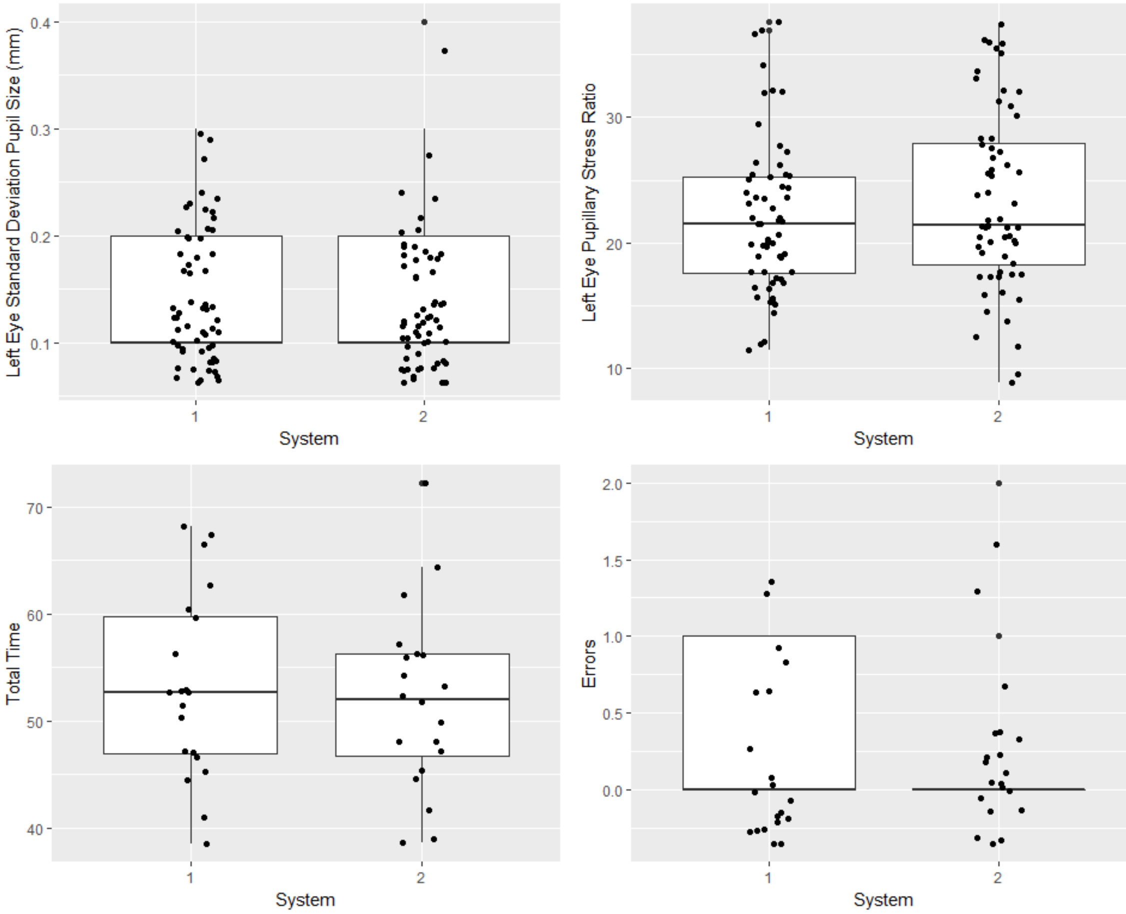
Appendix B. Subject-specific agreement charts




Standard Deviation Fixation Polyarea (mm2)
2000
1500
1000
500
1
2
System
Average Inter Fixation Time (ms)
400
350
300
250
200
1
2
System
Standard Deviation Inter Fixation Time (ms)
400
300
200
100
1
2
System
Saccade Fixation Ratio
2.1
1.8
1.5
1.2
1
2
System
Total Saccades (#)
180
150
120
90
60
1
2
System
Total Saccade Length (mm)
40000
30000
1
2
System


Total Only Saccade Length (mm)
15000
20000
25000
30000
1
2
System
Average Saccade Length (mm)
200
240
280
320
1
2
System
Standard Deviation Saccade Length (mm)
100
150
200
250
300
1
2
System
Total Only Saccade Length In Degrees (deg)
150
200
250
300
1
2
System
Average Saccade Length In Degrees (deg)
1.5
2.0
2.5
3.0
1
2
System
Standard Deviation Saccade Length In Degrees (deg)
1.0
1.5
2.0
2.5
3.0
1
2
System


Total Blinks
0
10
20
30
1
2
System
Blink Rate (blinks per minute)
0
25
50
75
100
1
2
System
Average Blink Duration (ms)
0
250
500
750
1
2
System
Standard Deviation Blink Duration (ms)
0
100
200
300
400
1
2
System
Average Inter Blink Duration (seconds)
0
5
10
15
1
2
System
Left Eye Average Pupil Size (mm)
2.4
2.8
3.2
3.6
1
2
System



Left Eye Standard Deviation Pupil Size (mm)
0.4
0.3
0.2
0.1
1
2
System
Left Eye Pupillary Stress Ratio
30
20
10
1
2
System
Total Time
70
60
50
40
1
2
System
Errors
2.0
1.5
1.0
0.5
0.0
1
2
System

Department of the Army
U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama 36362-0577
www.usaarl.army.mil

U.S. Army Medical Research and Materiel Command