MARCO M3/NIDEK TONOREF II AUTOREFRACTOR INTERMETHOD RELIABILITY STUDY

by

Joshua Smith-Hanen

This paper is submitted in partial fulfillment of the requirements for the degree of

Doctor of Optometry

Ferris State University
Michigan College of Optometry

May 2013

MARCO M3/NIDEK TONOREF II AUTOREFRACTOR
INTERMETHOD RELIABILITY STUDY

by

Joshua Smith-Hanen

Has been approved

May 2013

APPROVED:

, Faculty Advisor

ACCEPTED:

Ferris State University
Doctor of Optometry Senior Paper
Library Approval and Release

MARCO M3/NIDEK TONOREF II AUTOREFRACTOR
INTERMETHOD RELIABILITY STUDY

I, ____Joshua Smith-Hanen____, hereby release this Paper as described above to Ferris State University with the understanding that it will be accessible to the general public. This release is required under the provisions of the Federal Privacy Act.

ABSTRACT

Background: Autorefractors are commonly used as a starting point for subjective refraction and accordingly the repeatability and accuracy of these devices are of great importance to efficiently and effectively assessing refractive error. This study analyzes the repeatability and accuracy of three different autorefractors which are the same manufacturer model.

Methods: The instruments used in this study were three Marco M3/Nidek Tonoref II. Each participant was autorefracted using each of these three devices by the same trained operator. Participants were free of ocular media opacities and without ocular or systemic disease. Data was then analyzed for inter-method reliability/test-retest reliability.

Results: Raw data was collected with the traditional sphere, cylinder, and axis components. This data was then translated into components amenable to statistical analysis: M, J0, J45, and P. Bland-Altman analysis for these components of the three instruments is reported. Repeated measures multivariate analysis is also presented.

Discussion: Bland-Altman analysis shows the measurements obtained on each of the three instruments are similar. Repeated measures multivariate analysis shows that there isn't a statistically significant difference between the measurements obtained on each of the three instruments.

Conclusion: Since Bland-Altman analysis shows the instruments produce similar measurements and repeated measures multivariate analysis shows the measurements differ by a statistically insignificant amount, the Marco M3/Nidek Tonoref II shows reliability between instruments.

ACKNOWLEDGEMENT

Of the involved faculty, staff, and students of the Michigan College of Optometry, especially, Robert S. Buckingham, O.D., Ph.D., F.A.A.O., my advisor, for support of this study.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

BACKGROUND

Autorefractors are commonly used for objective refraction and as a starting point for subjective refraction. Accordingly, the repeatability and accuracy of these devices are of great importance to efficiently and effectively assessing refractive error. This study analyzes the repeatability and accuracy of three different instruments of the same manufacturer and model of autorefractor.

There are several previous studies relating to repeatability and accuracy of autorefraction. Several of these studies are meaningful for comparison to the present study. In the paragraphs below, these studies are described briefly and then distinguished from the present study.

One study assessed the repeatability of the Nikon NRK-8000, the Nidek AR-1000, and subjective refraction. In this study, measurements were taken with all three techniques on two separate occasions with a test-retest separation of at least 24 hours on only the right eyes of thirty normal subjects and found subjective refraction to be the most repeatable method.[1]

Another study examined the repeatability and validity of the PowerRefractor and the Nidek AR-600A autorefractor involving fifty subjects aged 16 to 61 years by comparing them to subjective refraction, and found no significant difference between the measurements obtained with the two instruments and the subjective refraction.[2]

A third study focused on comparing the repeatability of the Hoya AR-570 and clinician refraction with eighty-six subjects, aged 11 to 60 years, and concluded automated refraction is more repeatable than subjective refraction.[3]

A tangentially related repeatability study assessed the repeatability and agreement of refractive error measurements (and the repeatability of axial length measurements) in patients after laser in situ keratomileusis. Focusing only on the results of the autorefraction repeatability, autorefraction measurements with the Grand Seiko and Humphrey autorefractors, and subjective refraction were compared for 40 previously myopic LASIK patients under non-cycloplegic and cycloplegic conditions on two separate occasions. The study concluded there were statistically and clinically significant differences between subjective refraction and the Humphrey autorefractor regardless of cycloplegic state. The study further found insignificant differences between the Grand Seiko autorefractor and subjective refraction under non-cycloplegic conditions and statistically significant, yet clinically insignificant differences under cycloplegic conditions.[4]

Another comparison reliability study involved a retrospective analysis of 97 participants aged 18 to 66 years. Essentially, it involved a correlation comparison based on visual acuity of the Ophthonix Z-View aberrometer and a Humphrey autorefractor with standard subjective refraction. The study determined the Z-view aberrometer provided better accuracy than the Humphrey autorefractor as a starting point for subjective refraction.[5]

A study involving 190 subjects using either the Nidek ARK-700A (Fremont, CA) or the Topcon KR-8000 (Paramus, NJ) and subjective refraction (masked to autorefraction) were compared in terms of spherical equivalent using Bland-Altman limits of agreement and astigmatic vector difference using median and 95th percentile. The study found both autorefractors had clinical equivalence despite statistically significant differences and a small advantage for the Nidek in avoiding large astigmatic errors.[6]

A comparison of the Nidek AR-1000 autorefractor to retinoscopy with cycloplegia was conducted. The autorefractor was found to better determine the axis and that neither the sphere nor cylinder component differed significantly between autorefraction under influence of cyclopentolate and retinoscopy with atropine cycloplegia. When cycloplegic agents were not employed, the autorefractor was found to be less accurate in terms of the sphere power component, which may be attributable to accommodation and that, when accommodative disorders are suspected, cycloplegic autorefraction would be more accurate.

None of these studies directly focused on how repeatable the measurements were with one single model of autorefractor, as the present study does. Although comparison of different autorefractor reliability is useful, focusing on a specific model autorefractor determines whether and to what extent a particular model and manufacturer of autorefractor is reliable from instrument to instrument.

Further, subjective refraction plays a different role than autorefractors do in clinical practice, as the latter is purely objective and the former obviously involves subjective determinations. The subjective determinations are variable and not necessarily optically exact. Another variation occurs in the method of refraction employed and potential variability in methods by the examiner performing the refraction. The present study focuses only on the reliability of autorefraction from instrument to instrument without comparison to subjective refraction or another autorefaction model.

Some of these studies employed cycloplegic agents, which would take accommodation out of play, and may increase the accuracy of autorefraction. However, common clinical practice employs autorefraction prior to the instillation of cycloplegic agents. The reliability assessment of a specific model of autorefractor on a participant who may have accommodative dysfunction may cause variability in measurement; however, this study will not exclude patients on the basis of some potential accommodative dysfunction and will not employ cycloplegic agents. Furthermore, taking readings from three instruments at essentially the same time on the same day should mean any potential accommodative component would be present when readings are obtained at each of the instruments.

The Marco M3/Nidek tonoref II provides objective refraction by projecting fine measurement beams on the fundus of the subject's eye and then computation is performed by capturing the reflected beams as a ring image to measure the refractive error. This instrument is capable of capturing keratometry readings and non-contact tonometry as well.[7] However, this present study focuses only on the autorefraction readings.

METHODS

The three autorefractors used in this study are the same model and year of manufacturer and were the Marco M3/Nidek Tonoref II (Tokyo, Japan). These autorefractors were in good working order and well maintained. Each measurement was obtained while in automated shooting mode and three-dimensional autotracking mode. The autorefraction settings were left as manufacturer presets. Each participant was instructed to look at the fixation target (balloon) through the measuring window without straining. If a participant's eyelashes were potentially impacting the reading, the participant was instructed to open their eyes wider, which allowed for more accurate readings. The readings were obtained on the same day and in the same order in succession with only the time it takes to properly position the participant at each of the three autorefractors in between readings. The data was collected at the Michigan College of Optometry, Ferris State University.

Participants were free of corneal irregularities, ocular media opacities, and ocular or systemic disease. Rigid contact lens wearers were excluded. Participants' age and gender were noted at the time of data collection.

Data was compiled and analyzed with Microsoft Excel (Redmond, WA). Data was analyzed for inter-method reliability/test-retest reliability. The average refractive error at each of the three instruments was compared using Fourier analysis, which is a method of analysis used in similar studies.[8-13] The following are the pertinent equations used for calculating the power components of the autorefractor measurements for each instrument via Fourier analysis.

$M = S+(C/2)$ where S = sphere, C = cylinder

$J0 = -(C/2) \cos(2*axis)$ where C = cylinder

$J45 = -(C/2) \sin(2*axis)$ where C = cylinder

$P = (M^2 + J0^2 + J45^2)^{(1/2)}$

M, J0, J45, and P were calculated using the average of three readings for the right and left eye of each of the three instruments.

RESULTS

Scatter plots and Bland-Altman plots were obtained comparing each of the instruments and both of the subject's eyes at each of these instruments. These plots comparing M for each eye and each of the three instruments are included as Appendix A. The J0 and J45 comparison analysis plots are included as Appendix B and C respectively. Comparison analysis plots for P are included as Appendix D. Summary tables of the values for the mean, standard deviation, and 1.96-standard deviation for the M, P, J0 and J45 for each eye are included as Appendix E.

Bland-Altman analysis shows that the measurements of refractive error obtained by each Marco M3/Nidek Tonoref II and analyzed via Fourier component parts are similar. The standard deviation does not exceed one-half of a diopter for any component of the measurements obtained. With the exception of a limited number of outliers, measurements fall within two standard deviations of the mean. Ninety-two percent or greater of M component measurements are within two standard deviations of the mean. Ninety-five percent of all other Fourier component measurements are within two standard deviations of the mean.

Multivariate analysis was utilized for comparing the M, J0, J45, and P. Multivariate analysis was utilized because this study compares refractive error measurements obtained on three different instruments of the same make and model on the same persons. Refractive error measurements are compared and are assessed for statistical significance in terms of M, J0, J45, and P. In addition, gender was evaluated for statistical significance.

The following is the multivariate analysis for M:

Table 1.

Descriptive Statistics

	Gender	Mean	Std. Deviation	N
M-1	Female	-2.21591	2.349599	110
	Male	-2.36574	2.435955	54
	Total	-2.26524	2.371942	164
M-2	Female	-2.25114	2.310321	110
	Male	-2.38426	2.411017	54
	Total	-2.29497	2.337383	164
M-3	Female	-2.23409	2.327052	110
	Male	-2.34259	2.445658	54
	Total	-2.26982	2.359793	164

Table 2.

Box's Test of Equality of Covariance Matrices[a]

Box's M	9.414
F	1.532
df1	6
df2	73528.885
Sig.	.163

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Gender
Within Subjects Design: M

Table 3.

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power[b]
M	Pillai's Trace	.022	1.844[a]	2.000	161.000	.161	3.689	.380
	Wilks' Lambda	.978	1.844[a]	2.000	161.000	.161	3.689	.380
	Hotelling's Trace	.023	1.844[a]	2.000	161.000	.161	3.689	.380
	Roy's Largest Root	.023	1.844[a]	2.000	161.000	.161	3.689	.380
M * Gender	Pillai's Trace	.007	.566[a]	2.000	161.000	.569	1.132	.143
	Wilks' Lambda	.993	.566[a]	2.000	161.000	.569	1.132	.143
	Hotelling's Trace	.007	.566[a]	2.000	161.000	.569	1.132	.143
	Roy's Largest Root	.007	.566[a]	2.000	161.000	.569	1.132	.143

a. Exact statistic

b. Computed using alpha = .05

c. Design: Intercept + Gender
Within Subjects Design: M

Table 4.

Mauchly's Test of Sphericity[b]

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
M	.938	10.296	2	.006	.942	.958	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + Gender
Within Subjects Design: M

Table 5.

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
M	Sphericity Assumed	.077	2	.038	1.586	.206	3.172	.335
	Greenhouse-Geisser	.077	1.883	.041	1.586	.208	2.987	.325
	Huynh-Feldt	.077	1.917	.040	1.586	.207	3.040	.328
	Lower-bound	.077	1.000	.077	1.586	.210	1.586	.240
M * Gender	Sphericity Assumed	.031	2	.016	.647	.524	1.294	.158
	Greenhouse-Geisser	.031	1.883	.017	.647	.515	1.218	.155
	Huynh-Feldt	.031	1.917	.016	.647	.518	1.240	.156
	Lower-bound	.031	1.000	.031	.647	.422	.647	.126
Error(M)	Sphericity Assumed	7.843	324	.024				
	Greenhouse-Geisser	7.843	305.099	.026				
	Huynh-Feldt	7.843	310.475	.025				
	Lower-bound	7.843	162.000	.048				

a. Computed using alpha = .05

Table 6.

Tests of Within-Subjects Contrasts

Source	M	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
M	Level 2 vs. Level 1	.105	1	.105	2.877	.092	2.877	.392
	Level 3 vs. Level 1	.001	1	.001	.016	.899	.016	.052
M * Gender	Level 2 vs. Level 1	.010	1	.010	.278	.599	.278	.082
	Level 3 vs. Level 1	.062	1	.062	1.130	.289	1.130	.184
Error(M)	Level 2 vs. Level 1	5.892	162	.036				
	Level 3 vs. Level 1	8.872	162	.055				

a. Computed using alpha = .05

Table 7.

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
M-1	.051	1	162	.821
M-2	.065	1	162	.799
M-3	.115	1	162	.735

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Gender
Within Subjects Design: M

Table 8.

Tests of Between-Subjects Effects

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	765.708	1	765.708	137.545	.000	137.545	1.000
Gender	.617	1	.617	.111	.740	.111	.063
Error	901.848	162	5.567				

a. Computed using alpha = .05

The Box's Test (table 2) tells us that the covariance of the dependent variables is not statistically different. Therefore, we use the Wilks' lambda in the multivariate test.[14,15] As shown in table 3, the Wilks' lambda tells us that there isn't a statistically significant difference between M-1, M-2, M-3, nor for gender. Maulchy's test evaluates whether the sphericity assumption has been violated. As shown in table 6, with a significance of 0.006, sphericity cannot be assumed. When episolon is greater than 0.75, the Huynh-Feldt correction should be applied.[16,17] The Huynh-Feldt correction tells us that for M, there isn't a statistically significant difference between M-1, M-2, and M-3. Further, Huynh-Feldt tells us that there isn't a statistically significant difference between M-1, M-2, and M-3 when comparing gender. As shown in table 6, there isn't a statistically significant difference between M-1 and M-2, M-3 and M-1, nor when gender is considered. Levene's test, as shown in table 7, tells us that the variance of the dependent variables is not statistically different. Also, there is not a statistically significant difference between gender, as shown in table 8.

The following is the analysis for J0:

Table 9

Descriptive Statistics

	Gender	Mean	Std. Deviation	N
J0 – 1	Female	.012872	.2106764	110
	Male	-.036389	.2672590	54
	Total	-.003348	.2311807	164
J0 – 2	Female	.021193	.1980739	110
	Male	.041541	.2829570	54
	Total	.027893	.2288254	164
J0 – 3	Female	.004457	.2166440	110
	Male	-.022766	.3373854	54
	Total	-.004507	.2618440	164

Table 10

Box's Test of Equality of Covariance Matrices[a]

Box's M	38.223
F	6.219
df1	6
df2	73528.885
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Gender
Within Subjects Design: J0

Table 11

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power[b]
J0	Pillai's Trace	.023	1.906[a]	2.000	161.000	.152	3.813	.392
	Wilks' Lambda	.977	1.906[a]	2.000	161.000	.152	3.813	.392
	Hotelling's Trace	.024	1.906[a]	2.000	161.000	.152	3.813	.392
	Roy's Largest Root	.024	1.906[a]	2.000	161.000	.152	3.813	.392
J0 * Gender	Pillai's Trace	.012	.984[a]	2.000	161.000	.376	1.967	.219
	Wilks' Lambda	.988	.984[a]	2.000	161.000	.376	1.967	.219
	Hotelling's Trace	.012	.984[a]	2.000	161.000	.376	1.967	.219
	Roy's Largest Root	.012	.984[a]	2.000	161.000	.376	1.967	.219

a. Exact statistic

b. Computed using alpha = .05

c. Design: Intercept + Gender
Within Subjects Design: J180

Table 12.

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
J0	.927	12.227	2	.002	.932	.948	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
b. Design: Intercept + Gender
Within Subjects Design: J0

Table 13

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
J0	Sphericity Assumed	.169	2	.085	1.404	.247	2.808	.301
	Greenhouse-Geisser	.169	1.864	.091	1.404	.247	2.616	.290
	Huynh-Feldt	.169	1.896	.089	1.404	.247	2.662	.293
	Lower-bound	.169	1.000	.169	1.404	.238	1.404	.218
J0 * Gender	Sphericity Assumed	.092	2	.046	.760	.469	1.519	.179
	Greenhouse-Geisser	.092	1.864	.049	.760	.460	1.416	.174
	Huynh-Feldt	.092	1.896	.048	.760	.462	1.440	.175
	Lower-bound	.092	1.000	.092	.760	.385	.760	.139
Error(J0)	Sphericity Assumed	19.553	324	.060				
	Greenhouse-Geisser	19.553	301.919	.065				
	Huynh-Feldt	19.553	307.180	.064				
	Lower-bound	19.553	162.000	.121				

a. Computed using alpha = .05

Table 14

Tests of Within-Subjects Contrasts

Source	J180	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
J0	Level 2 vs. Level 1	.269	1	.269	2.535	.113	2.535	.353
	Level 3 vs. Level 1	.001	1	.001	.006	.936	.006	.051
J0 * Gender	Level 2 vs. Level 1	.176	1	.176	1.651	.201	1.651	.248
	Level 3 vs. Level 1	.018	1	.018	.115	.735	.115	.063
Error(J0)	Level 2 vs. Level 1	17.216	162	.106				
	Level 3 vs. Level 1	24.829	162	.153				

a. Computed using alpha = .05

Table 15

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
J0 - 1	1.876	1	162	.173
J0 - 2	2.197	1	162	.140
J0 - 3	3.274	1	162	.072

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept + Gender
Within Subjects Design: J0

Table 16

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	.002	1	.002	.098	.755	.098	.061
Gender	.013	1	.013	.705	.402	.705	.133
Error	2.913	162	.018				

a. Computed using alpha = .05

The Box's Test, as shown in table 10, tells us that the covariance of the dependent variables are statistically different. So, we should use Pillai's trace value.[14,15] As shown in table 11, Pillai's trace value reveals that for J0, there isn't a statistically significance difference between J0-1, J0-2, or J0-3, nor for gender. Again, Maulchy's test for sphericity evaluates whether the sphericity assumption has been violated. With a significance of 0.002, sphericity cannot be assumed. When epsilon is greater than 0.75, the Huynh-Feldt correction should be applied.[16,17] The Huynh-Feldt correction tells us that there isn't a statistically significant difference between J0-1, J0-2, J0-3. Further, Huynh-Feldt tells us there isn't a statistically significant difference between J0-1, J0-2, and J0-3 when comparing gender. Tests of within-subject contrasts, as shown in table 14, tells us there isn't a statistically significant difference between J0-1 and J0-2, J0-1 and J0-3, nor if gender is included. Levene's test of equality of error variance tells us that the variances of the dependent variables are not statistically different. Table 16 confirms via tests of between-subjects effects that there isn't a statistically significant difference between genders.

The analysis for J45 is as follows:

Table 17

Descriptive Statistics

	Gender	Mean	Std. Deviation	N
J45 -1	Female	.032782	.2065587	110
	Male	-.000093	.3413097	54
	Total	.021957	.2581659	164
J45 -2	Female	-.016459	.2155906	110
	Male	.010829	.3285058	54
	Total	-.007474	.2575578	164
J45 -3	Female	-.035053	.2187439	110
	Male	.025832	.3051229	54
	Total	-.015006	.2511824	164

Table 18

Box's Test of Equality of Covariance Matrices[a]

Box's M	48.274
F	7.854
df1	6
df2	73528.885
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Gender
Within Subjects Design: J45

Table 19

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power[b]
J45	Pillai's Trace	.004	.337[a]	2.000	161.000	.714	.674	.103
	Wilks' Lambda	.996	.337[a]	2.000	161.000	.714	.674	.103
	Hotelling's Trace	.004	.337[a]	2.000	161.000	.714	.674	.103
	Roy's Largest Root	.004	.337[a]	2.000	161.000	.714	.674	.103
J45 * Gender	Pillai's Trace	.015	1.253[a]	2.000	161.000	.288	2.507	.270
	Wilks' Lambda	.985	1.253[a]	2.000	161.000	.288	2.507	.270
	Hotelling's Trace	.016	1.253[a]	2.000	161.000	.288	2.507	.270
	Roy's Largest Root	.016	1.253[a]	2.000	161.000	.288	2.507	.270

a. Exact statistic

b. Computed using alpha = .05

c. Design: Intercept + Gender
Within Subjects Design: J45

Table 20

Mauchly's Test of Sphericity[b]

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
J45	.950	8.254	2	.016	.952	.969	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
b. Design: Intercept + Gender
Within Subjects Design: J45

Table 21

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
J45	Sphericity Assumed	.039	2	.020	.347	.707	.694	.105
	Greenhouse-Geisser	.039	1.905	.021	.347	.697	.661	.104
	Huynh-Feldt	.039	1.939	.020	.347	.700	.673	.105
	Lower-bound	.039	1.000	.039	.347	.557	.347	.090
J45 * Gender	Sphericity Assumed	.163	2	.082	1.451	.236	2.902	.310
	Greenhouse-Geisser	.163	1.905	.086	1.451	.236	2.764	.302
	Huynh-Feldt	.163	1.939	.084	1.451	.236	2.813	.305
	Lower-bound	.163	1.000	.163	1.451	.230	1.451	.224
Error(J45)	Sphericity Assumed	18.252	324	.056				
	Greenhouse-Geisser	18.252	308.579	.059				
	Huynh-Feldt	18.252	314.082	.058				
	Lower-bound	18.252	162.000	.113				

a. Computed using alpha = .05

Table 22

Tests of Within-Subjects Contrasts

Source	J45	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
J45	Level 2 vs. Level 1	.053	1	.053	.570	.451	.570	.117
	Level 3 vs. Level 1	.064	1	.064	.467	.495	.467	.104
J45 * Gender	Level 2 vs. Level 1	.131	1	.131	1.405	.238	1.405	.218
	Level 3 vs. Level 1	.318	1	.318	2.336	.128	2.336	.330
Error(J45)	Level 2 vs. Level 1	15.115	162	.093				
	Level 3 vs. Level 1	22.080	162	.136				

a. Computed using alpha = .05

Table 23

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
J45 -1	4.028	1	162	.046
J45 -2	2.579	1	162	.110
J45 -3	3.458	1	162	.065

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Gender
Within Subjects Design: J45

Table 24

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	.001	1	.001	.046	.830	.046	.055
Gender	.012	1	.012	.443	.507	.443	.101
Error	4.503	162	.028				

a. Computed using alpha = .05

As with J0, the Box's Test tells us that the covariance of dependent variables is statistically different so we should use the Pillai's trace value, which is shown in table 18.[14,15] Pillai's trace reveals to us that for J45, there isn't a statistically significant difference between J45-1, J45-2, J45-3, nor for gender. Maulchy's test evaluates whether the sphericity assumption has been violated. Sphericity cannot be assumed with a significance of 0.016. When epsilon is greater than 0.75, as it is here, the Huynh-Feldt correction should again be applied.[16,17] The Huynh-Feldt tells us that for J45 there isn't a statistically significant difference between J45-1, J45-2, and J45-3. The Huynh-Feldt tells us that for J45 when comparing gender there isn't a statistically significant difference between J45-1, J45-2, and J45-3. Tests of within subjects-contrasts reveals that there isn't a statistically significant difference between J45-1 and J45-2 nor J45-1 and J45-3 nor if you include gender. Levene's test tells us that the variance of the dependent variables are not statistically different. Tests of between- subjects effects confirms that there isn't a statistically significant difference between gender.

The analysis for P is the following:

Table 25

Descriptive Statistics

	Gender	Mean	Std. Deviation	N
P –1	Female	2.397182	2.1828921	110
	Male	2.509620	2.3256068	54
	Total	2.434204	2.2243667	164
P –2	Female	2.406606	2.1663897	110
	Male	2.503646	2.3255750	54
	Total	2.438558	2.2133790	164
P –3	Female	2.396684	2.1797648	110
	Male	2.511502	2.3138725	54
	Total	2.434490	2.2183557	164

Table 26

Box's Test of Equality of Covariance Matrices[a]

Box's M	9.048
F	1.472
df1	6
df2	73528.885
Sig.	.183

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Gender
Within Subjects Design: P

Table 27

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Noncent. Parameter	Observed Power[b]
P	Pillai's Trace	.000	.007[a]	2.000	161.000	.993	.014	.051
	Wilks' Lambda	1.000	.007[a]	2.000	161.000	.993	.014	.051
	Hotelling's Trace	.000	.007[a]	2.000	161.000	.993	.014	.051
	Roy's Largest Root	.000	.007[a]	2.000	161.000	.993	.014	.051
P * Gender	Pillai's Trace	.002	.195[a]	2.000	161.000	.823	.389	.080
	Wilks' Lambda	.998	.195[a]	2.000	161.000	.823	.389	.080
	Hotelling's Trace	.002	.195[a]	2.000	161.000	.823	.389	.080
	Roy's Largest Root	.002	.195[a]	2.000	161.000	.823	.389	.080

a. Exact statistic

b. Computed using alpha = .05

c. Design: Intercept + Gender
Within Subjects Design: P

Table 28

Mauchly's Test of Sphericity[b]

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
P	.934	10.946	2	.004	.938	.955	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
b. Design: Intercept + Gender
Within Subjects Design: P

Table 29

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
P	Sphericity Assumed	.000	2	.000	.005	.995	.011	.051
	Greenhouse-Geisser	.000	1.877	.000	.005	.993	.010	.051
	Huynh-Feldt	.000	1.910	.000	.005	.993	.010	.051
	Lower-bound	.000	1.000	.000	.005	.942	.005	.051
P * Gender	Sphericity Assumed	.007	2	.003	.166	.847	.333	.076
	Greenhouse-Geisser	.007	1.877	.004	.166	.833	.312	.075
	Huynh-Feldt	.007	1.910	.004	.166	.837	.318	.075
	Lower-bound	.007	1.000	.007	.166	.684	.166	.069
Error(P)	Sphericity Assumed	6.565	324	.020				
	Greenhouse-Geisser	6.565	304.017	.022				
	Huynh-Feldt	6.565	309.354	.021				
	Lower-bound	6.565	162.000	.041				

a. Computed using alpha = .05

Table 30

Tests of Within-Subjects Contrasts

Source	P	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
P	Level 2 vs. Level 1	.000	1	.000	.014	.905	.014	.052
	Level 3 vs. Level 1	.000	1	.000	.001	.969	.001	.050
P * Gender	Level 2 vs. Level 1	.009	1	.009	.284	.595	.284	.083
	Level 3 vs. Level 1	.000	1	.000	.004	.947	.004	.050
Error(P)	Level 2 vs. Level 1	4.894	162	.030				
	Level 3 vs. Level 1	7.575	162	.047				

a. Computed using alpha = .05

Table 31

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
P -1	.334	1	162	.564
P -2	.368	1	162	.545
P -3	.259	1	162	.611

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Gender
Within Subjects Design: P

Table 32

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[a]
Intercept	872.619	1	872.619	176.754	.000	176.754	1.000
Gender	.423	1	.423	.086	.770	.086	.060
Error	799.778	162	4.937				

a. Computed using alpha = .05

The Box's Test tells us that the covariance of the dependent variables is not statistically different. Accordingly, we use the Wilks' lambda in the multivariate test.[14,15] The Wilks' lambda tells us that for P there isn't a statistically significant difference between P-1, P-2, P-3, nor for gender. Mualchy's test tells us whether the sphericity assumption has been violated. With a significance of 0.004, sphericity cannot be assumed. When epsilon is greater than 0.75, the Huynh-Feldt correction should be applied.[16,17] The Huynh-Feldt correction tells us that for P, there isn't a statistically significant difference between P-1, P-2, and P-3. The Huynh-Feldt correction for P gender tell us that there isn't a statistically significant difference between P-1, P-2, and P-3. Tests of within-subjects contrast revealed that there isn't a statistically significance difference P-1 and P-2 nor P-3 and P-1 nor if gender is included. Levene's test tells us that the variance of the dependent variables is not statistically different. Test of between-subjects effects reveals that there isn't a statistically significant difference between gender.

The following table summarizes the repeated measures analysis of each of these components.

Table 33.

Test	Repeated Measures Result	Conclusion
M	[F(2,310) = 1.586, p < 0.207]	There isn't a statistically significant difference between M-1, M-2, and M-3
J0	[F(2,307) = 1.404, p < 0.247]	There isn't a statistically significant difference between J0-1, J0-2, and J0-3
J45	[F(2,314) = 0.347, p < 0.700]	There isn't a statistically significant difference between J45-1, J45-2, and J45-3
P	[F(2,309) = 0.005, p < 0.993]	There isn't a statistically significant difference between P-1, P-2, and P-3

DISCUSSION

Bland-Altman analysis of data collected on each eye of 82 subjects showed that the measurements from instrument to instrument are similar. The measurements obtained on the three Marco M3/Nidek Tonoref II instruments, with the exception of a few outliers, all fall within 2 standard deviations of the mean. Therefore, the autorefractor measurements obtained with a Marco M3/Nidek Tonoref II are similar from instrument to instrument.

Repeated measures multivariate analysis revealed there isn't a statistically significant difference in the measurements obtained by three different Marco M3/Nidek Tonoref II instruments on each eye of 82 subjects. There is no statistically significant difference in terms of M, J0, J45, or P components of the refractive error measurements obtained. Therefore, the difference in autorefractor measurements obtained with a Marco M3/Nidek Tonoref II is statistically insignificant.

Accordingly, the Marco M3/Nidek Tonoref II provides measurements of refractive error which are similar among instruments and that differ by a statistically insignificant amount.

Being similar in measurement and differing in a statistically insignificant amount from instrument to instrument, means the Marco M3/Nidek Tonoref II has inter-instrument reliability and test-retest reliability. This reliability tends to suggest that the instrument is clinically reliable and provides a good starting point in the assessment of refractive error.

CONCLUSION

The accuracy of measurements provided by autorefractors is important as these devices provide objective refraction data and are commonly used as a starting point for subjective refraction. This study analyzed the repeatability and accuracy of three different Marco M3/Nidek Tonoref II autorefractors. Measurements were obtained by the same trained operator and participants were free of ocular media opacities, without ocular or systemic disease, and not rigid gas permeable contact lens wearers. Raw data was collected with the traditional sphere, cylinder, and axis components. This data was then translated into components amenable to statistical analysis via Fourier analysis: M, J0, J45, and P. Then, Bland-Altman analysis for these components of the three instruments was conducted and resulted in the conclusion that measurements between instruments are similar. Repeated measures multivariate analysis showed the measurements between instruments differ by a statistically insignificant amount when comparing any component or for gender. Therefore, Marco M3/Nidek Tonoref II instruments obtained measurements which are similar between instruments and that differ by a statistically insignificant amount between instruments. Accordingly, Marco M3/Nidek Tonoref II instruments show intermethod reliability and test-retest realiability.

REFERENCES

1. Elliott M, Simpson T, Richter D, Fonn D. Repeatability and accuracy of automated refraction: a comparison of the Nikon NRK-8000, the Nidek AR-1000, and subjective refraction. Optom Vis Sci. 1997 Jun;74(6):434–8.

2. Allen PM, Radhakrishnan H, O'Leary DJ. Repeatability and validity of the PowerRefractor and the Nidek AR600-A in an adult population with healthy eyes. Optom Vis Sci. 2003 Mar;80(3):245–51.

3. Bullimore MA, Fusaro RE, Adams CW. The repeatability of automated and clinician refraction. Optom Vis Sci. 1998 Aug;75(8):617–22.

4. Bailey MD, Twa MD, Mitchell GL, Dhaliwal DK, Jones LA, McMahon TT. Repeatability of autorefraction and axial length measurements after laser in situ keratomileusis. J Cataract Refract Surg. 2005 May;31(5):1025–34.

5. Cooper J, Citek K, Feldman JM. Comparison of refractive error measurements in adults with Z-View aberrometer, Humphrey autorefractor, and subjective refraction. Optometry - Journal of the American Optometric Association. 2011 Apr;82(4):231–40.

6. Pesudovs K, Weisinger HS. A comparison of autorefractor performance. Optom Vis Sci. 2004 Jul;81(7):554–8.

7. Nidek Co. Ltd. Tonoref II Operator's Manual, Tokyo, Japan.

8. De Juan V, Herreras JM, Martin R, Morejon A, Perez I, Cristobal AR-S, et al. Repeatability and agreement of ARK-30 autorefraction after cataract surgery. Clinical & Experimental Ophthalmology. 2012;40(2):134–40.

9. Groth AD, Hollingsworth SR, Ofri R, Kass PH, Reed Z, Murphy CJ. Clinical comparison of the Welch Allyn SureSight™ handheld autorefractor vs. streak retinoscopy in dogs. Vet Ophthalmol. 2012 Nov 22;1-5.

10. Choong Y-F, Chen A-H, Goh P-P. A comparison of autorefraction and subjective refraction with and without cycloplegia in primary school children. Am. J. Ophthalmol. 2006 Jul;142(1):68–74.

11. Atchison DA. Comparison of peripheral refractions determined by different instruments. Optom Vis Sci. 2003 Sep;80(9):655–60.

12. Gwiazda J, Weber C. Comparison of spherical equivalent refraction and astigmatism measured with three different models of autorefractors. Optom Vis Sci. 2004 Jan;81(1):56–61.

13. Pesudovs K, Weisinger HS. A comparison of autorefractor performance. Optom Vis Sci. 2004 Jul;81(7):554–8.

14. Meyers LS, Gamst G, Guarino AJ. Applied Multivariate Research: Design and Interpretation. SAGE; 2012, 232-241.

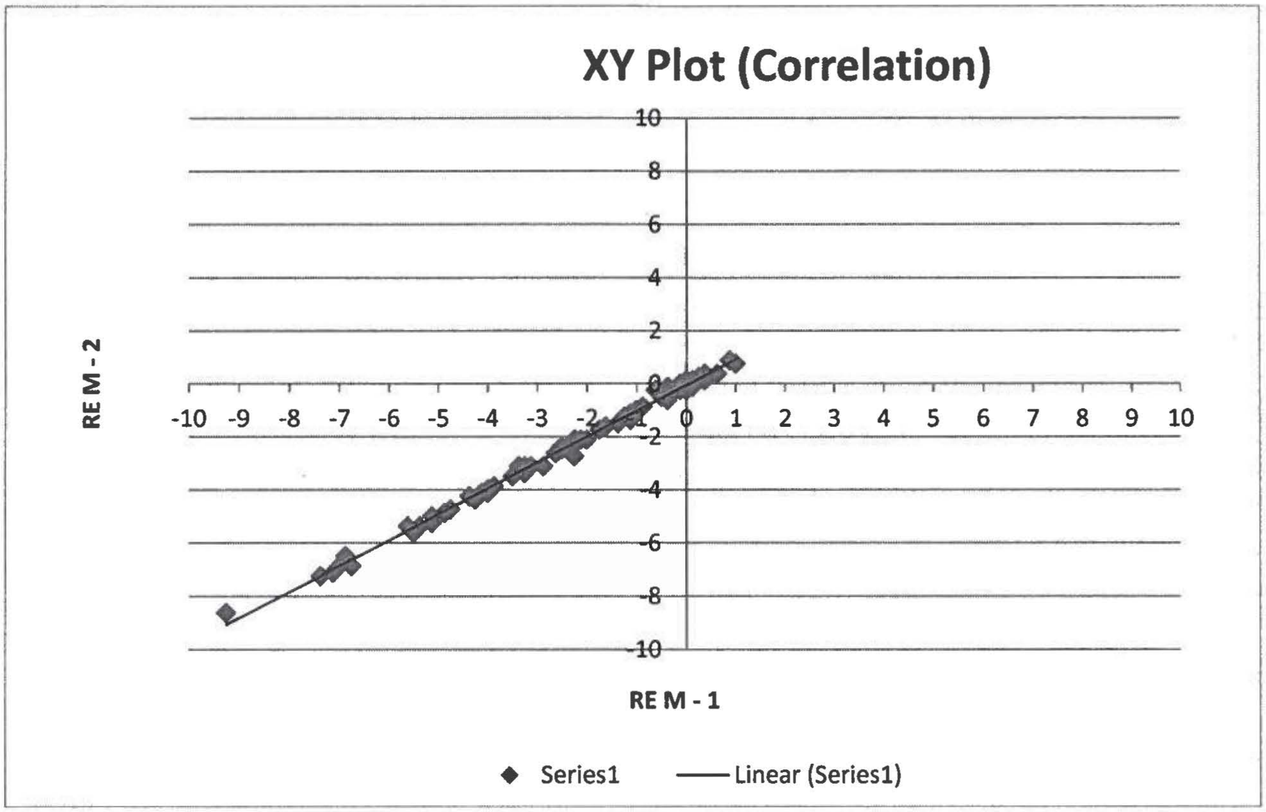
15. Raykov T, Marcoulides GA. An Introduction to Applied Multivariate Analysis. CRC Press; 2008, 193-209.

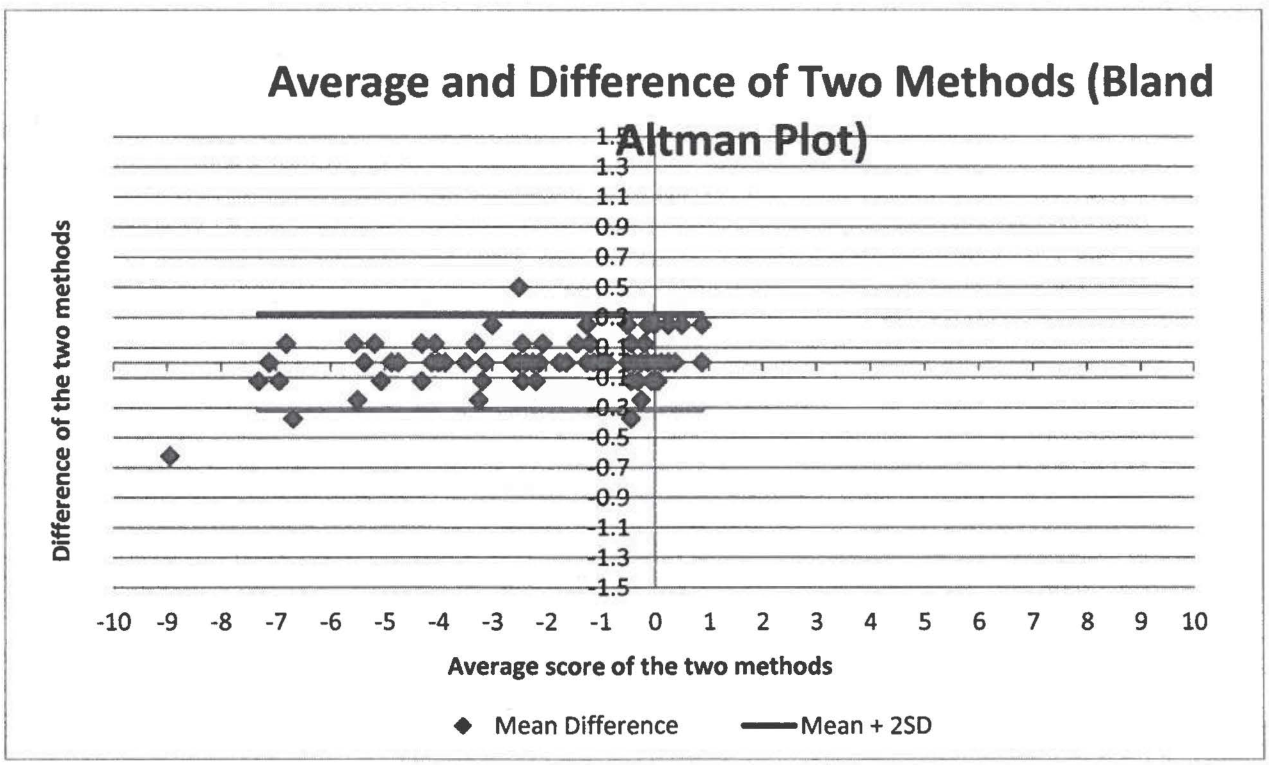
16. Kim M. Statistics for Evidence-Based Practice in Nursing. Jones & Bartlett Publishers; 2013, 294-301.

17. Quinn GGP, Keough MJ. Experimental Design and Data Analysis for Biologists. Cambridge University Press; 2002, 315-320.

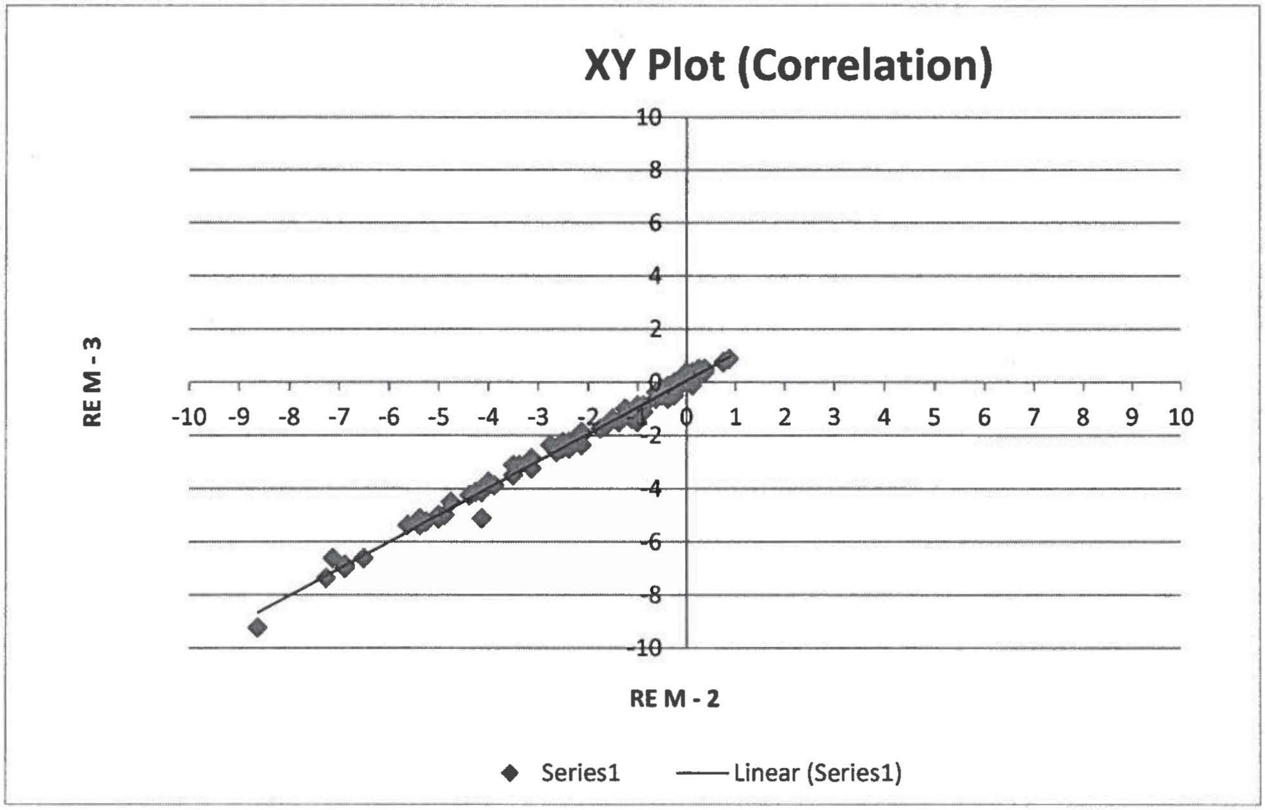
Appendix A

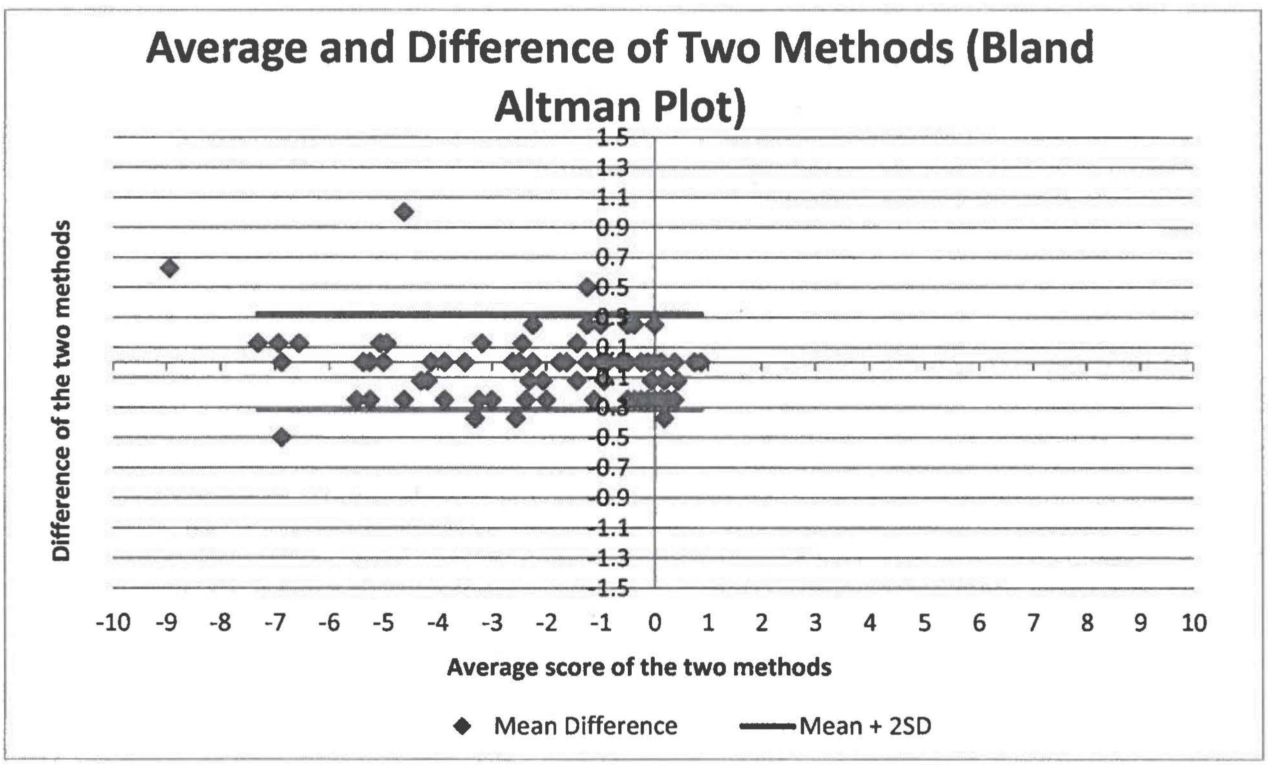
Appendix A: Right eye, M-value for Instrument 1 versus Instrument 2



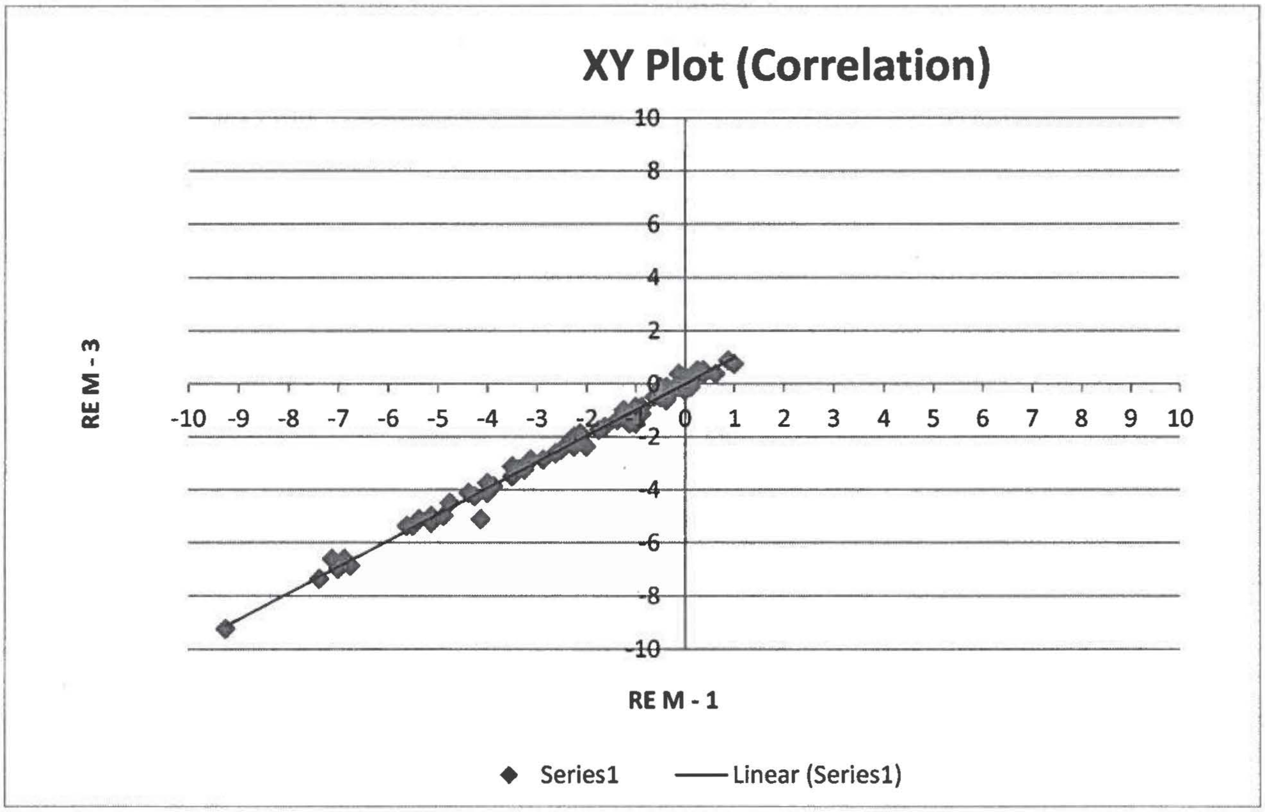


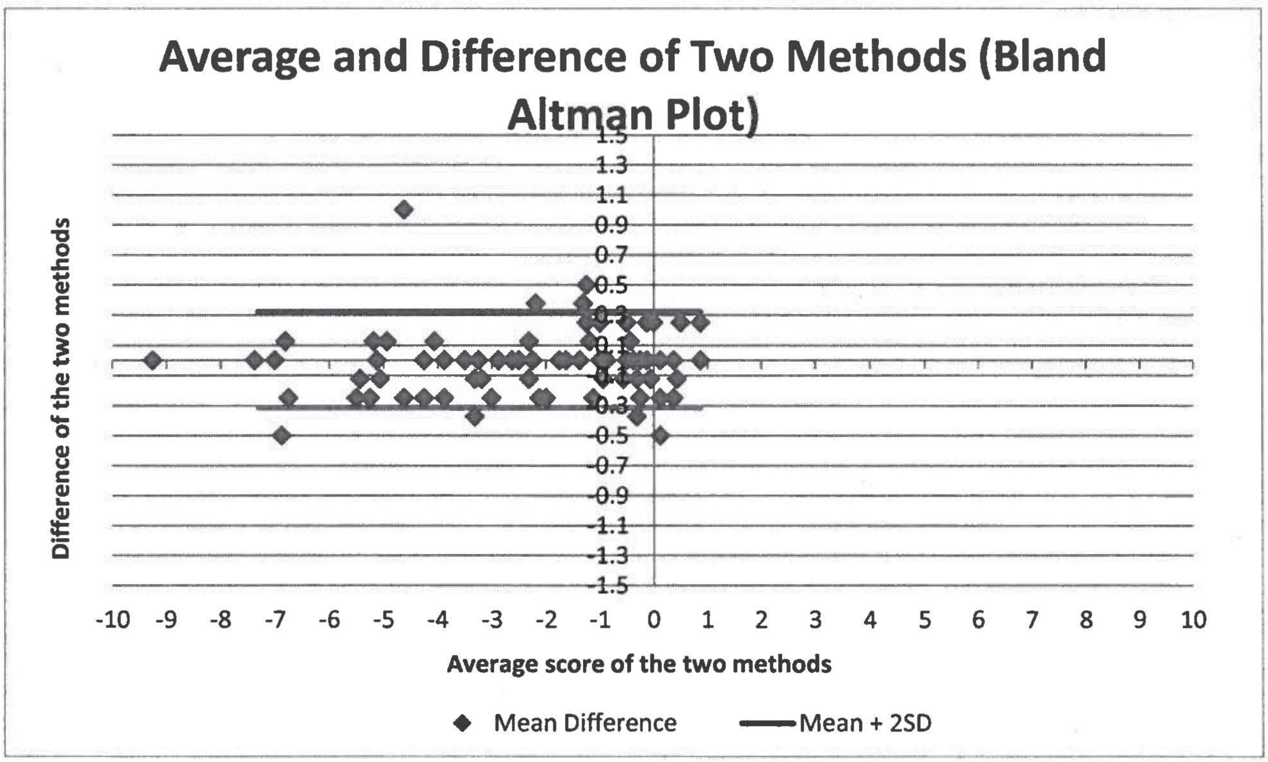
Appendix A: Right eye, M-value for Instrument 2 versus Instrument 3



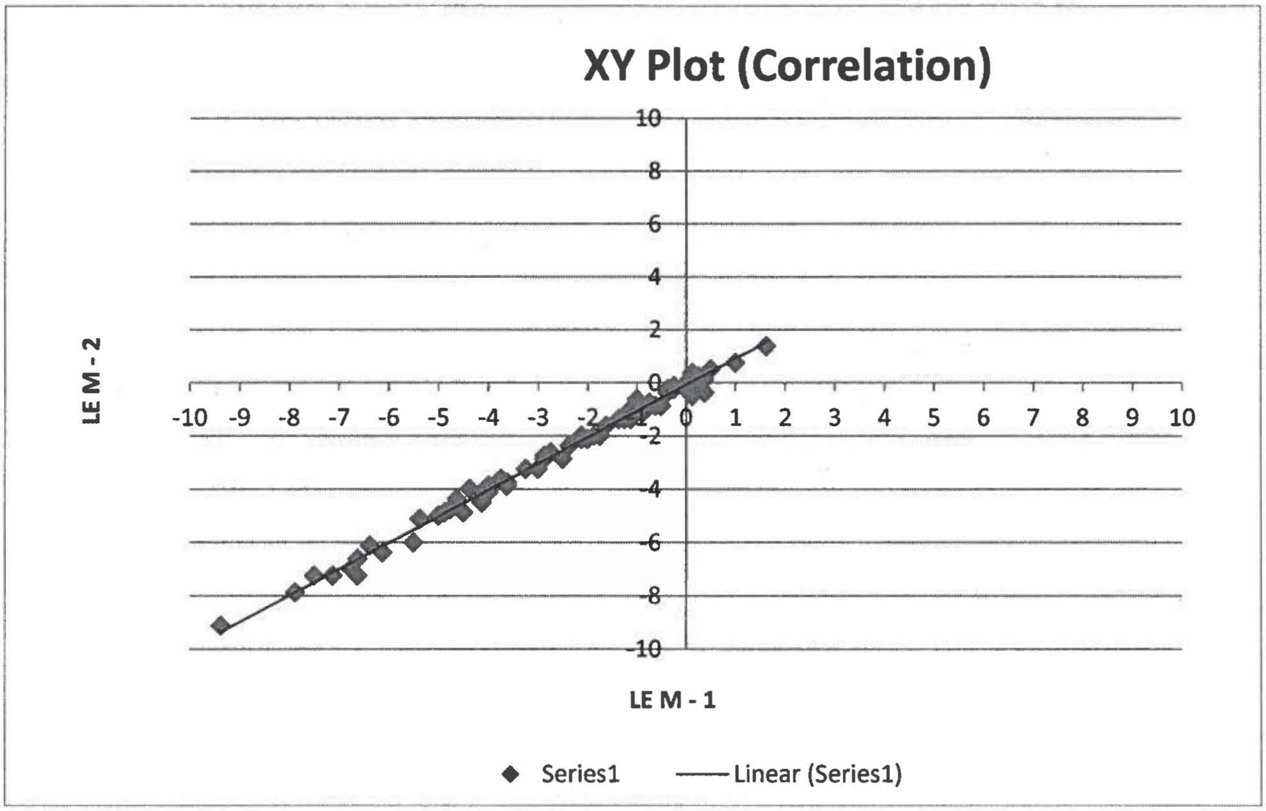


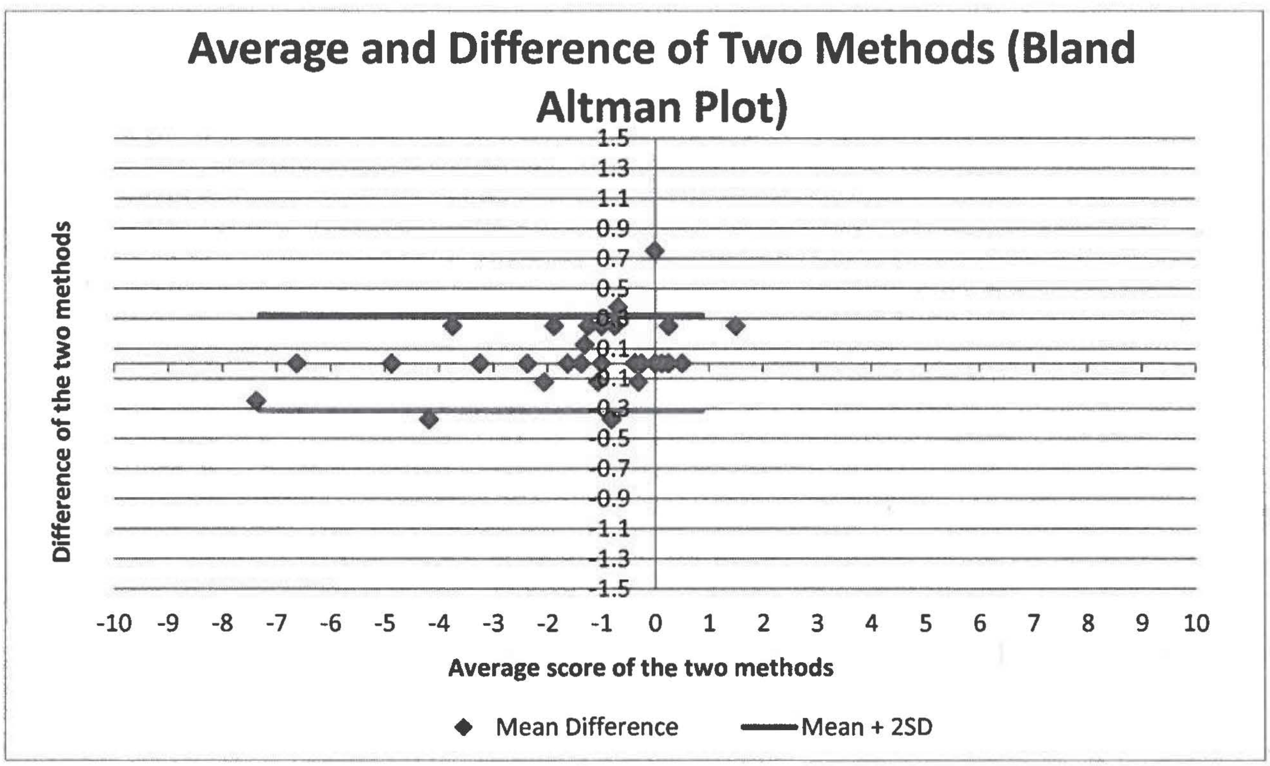
Appendix A: Right eye, M-value for Instrument 1 versus Instrument 3



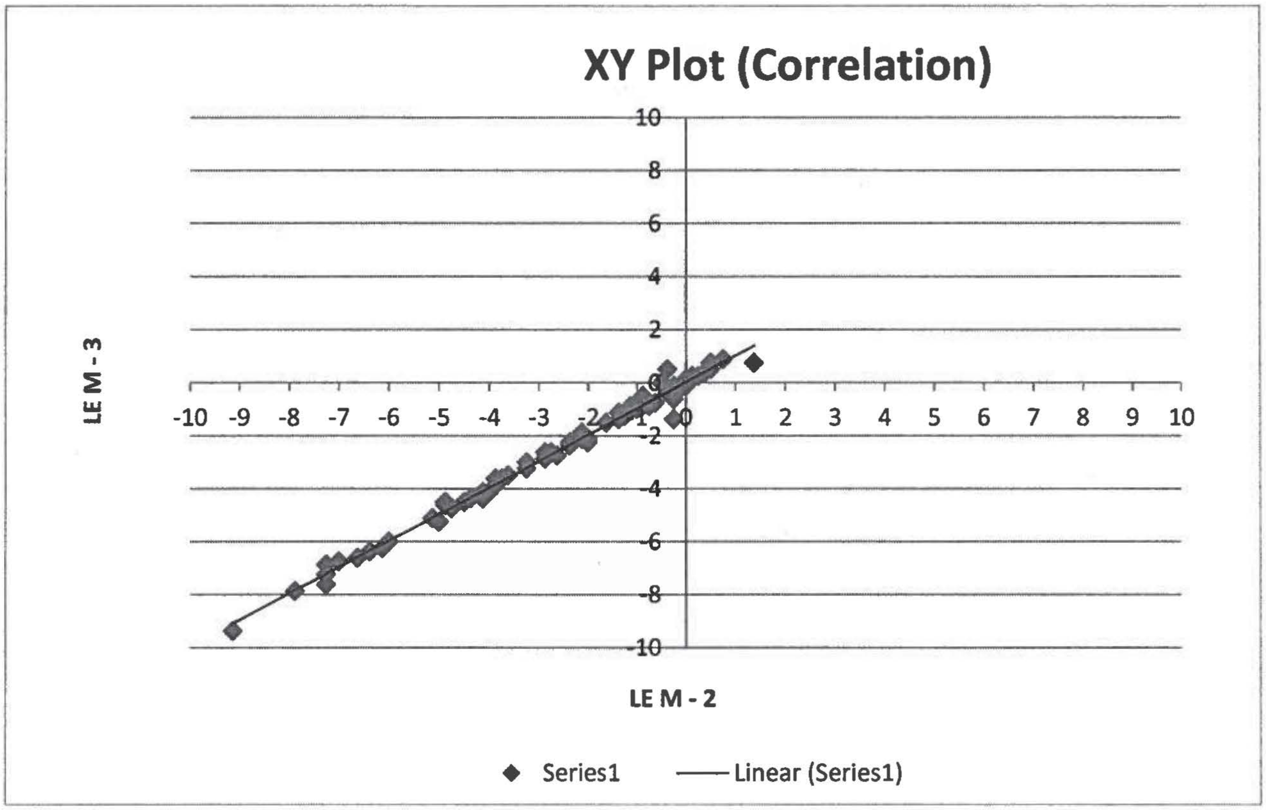


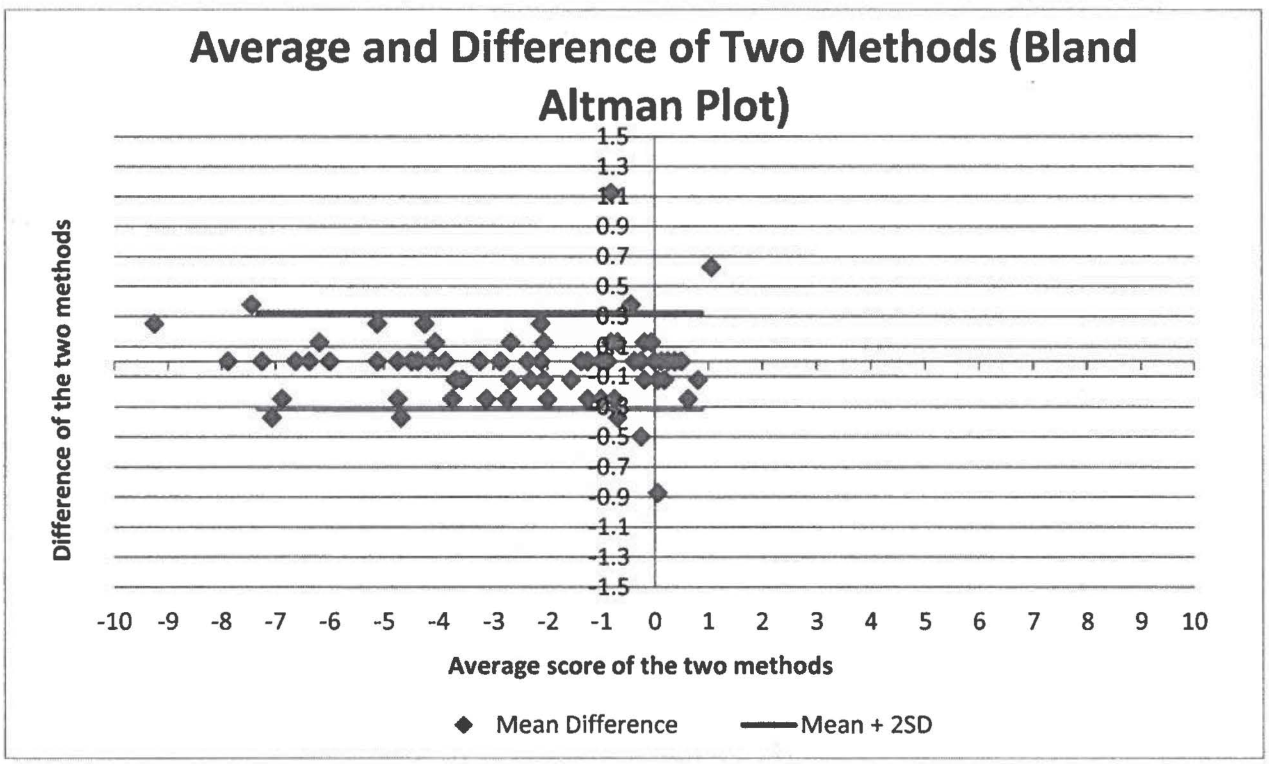
Appendix A: Left eye, M-value for Instrument 1 versus Instrument 3



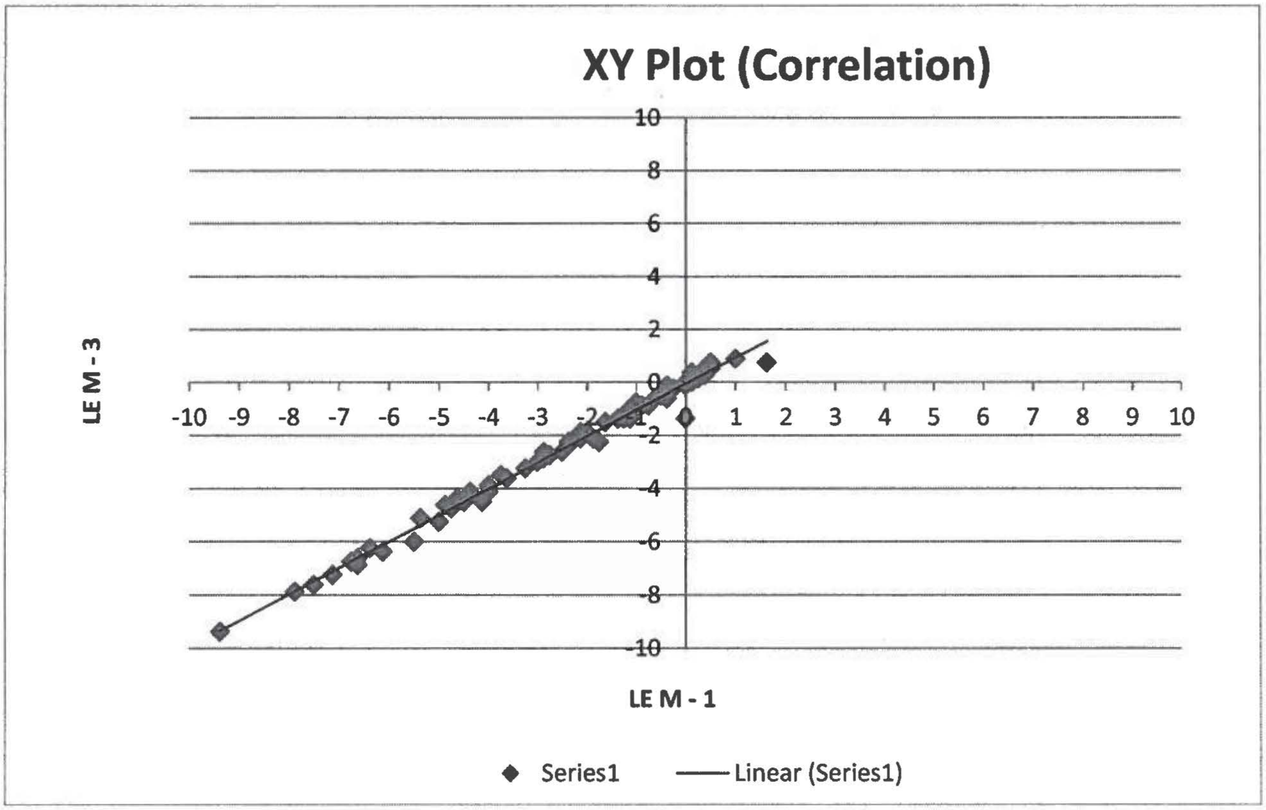


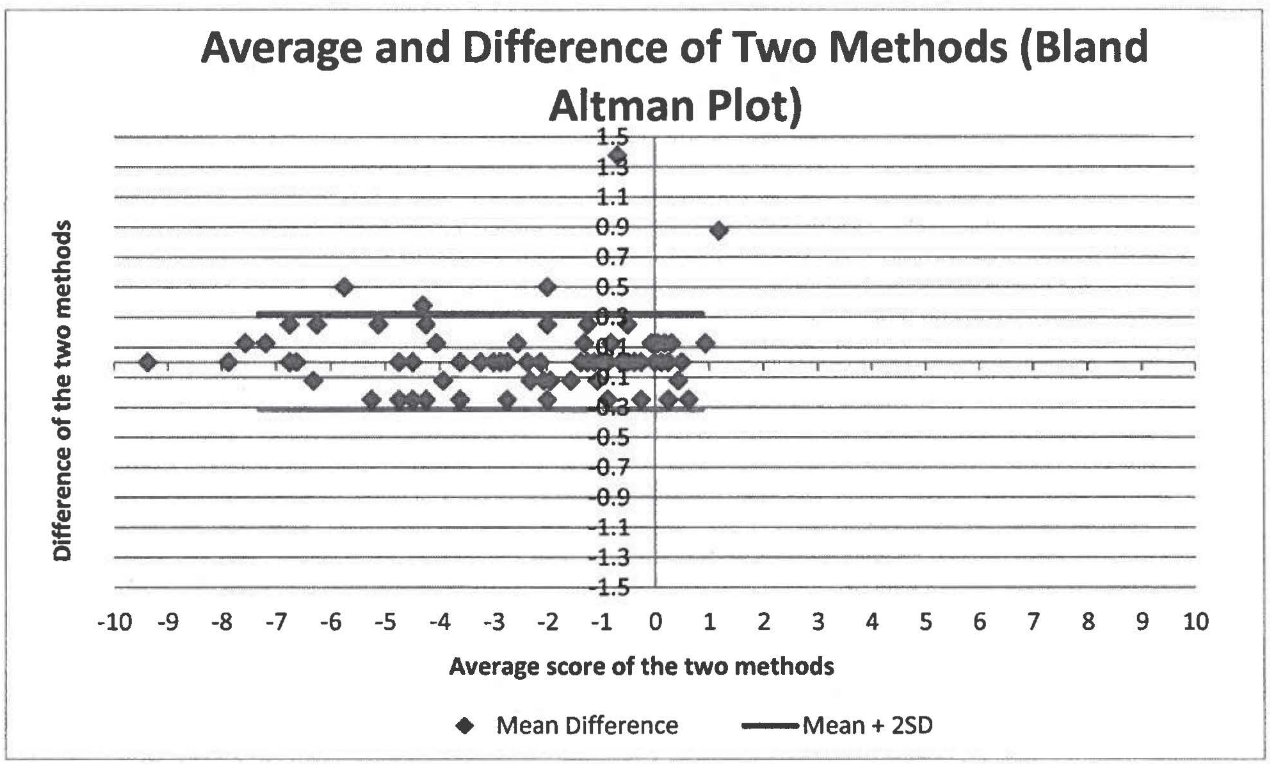
Appendix A: Left eye, M-value for Instrument 2 versus Instrument 3





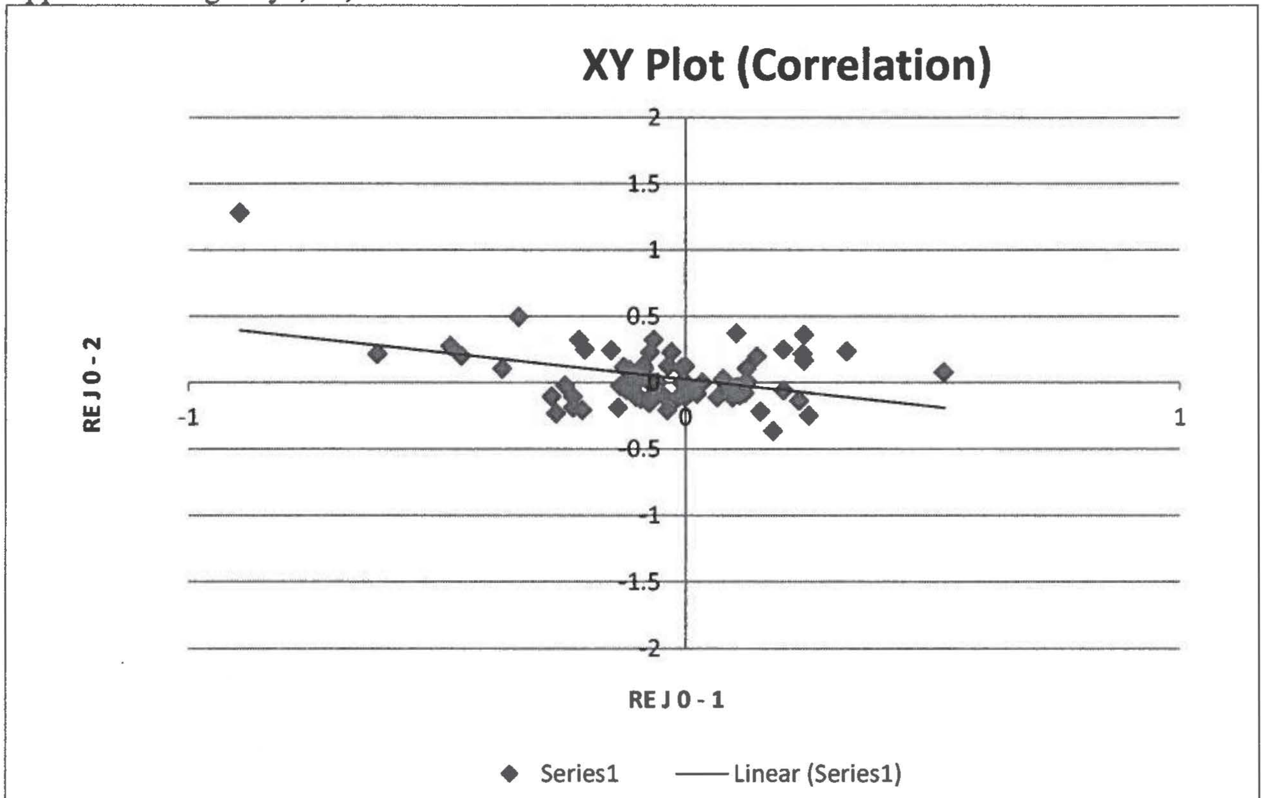
Appendix A: Left eye, M-value for Instrument 1 versus Instrument 3

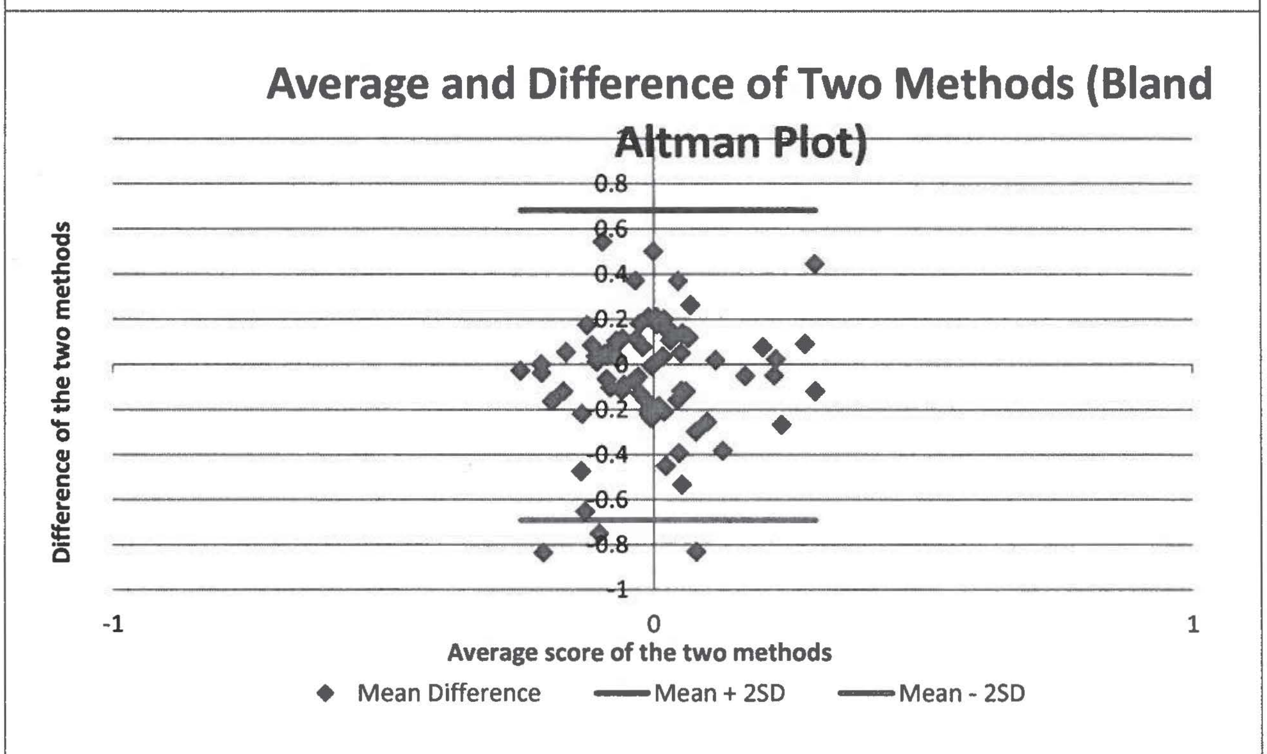




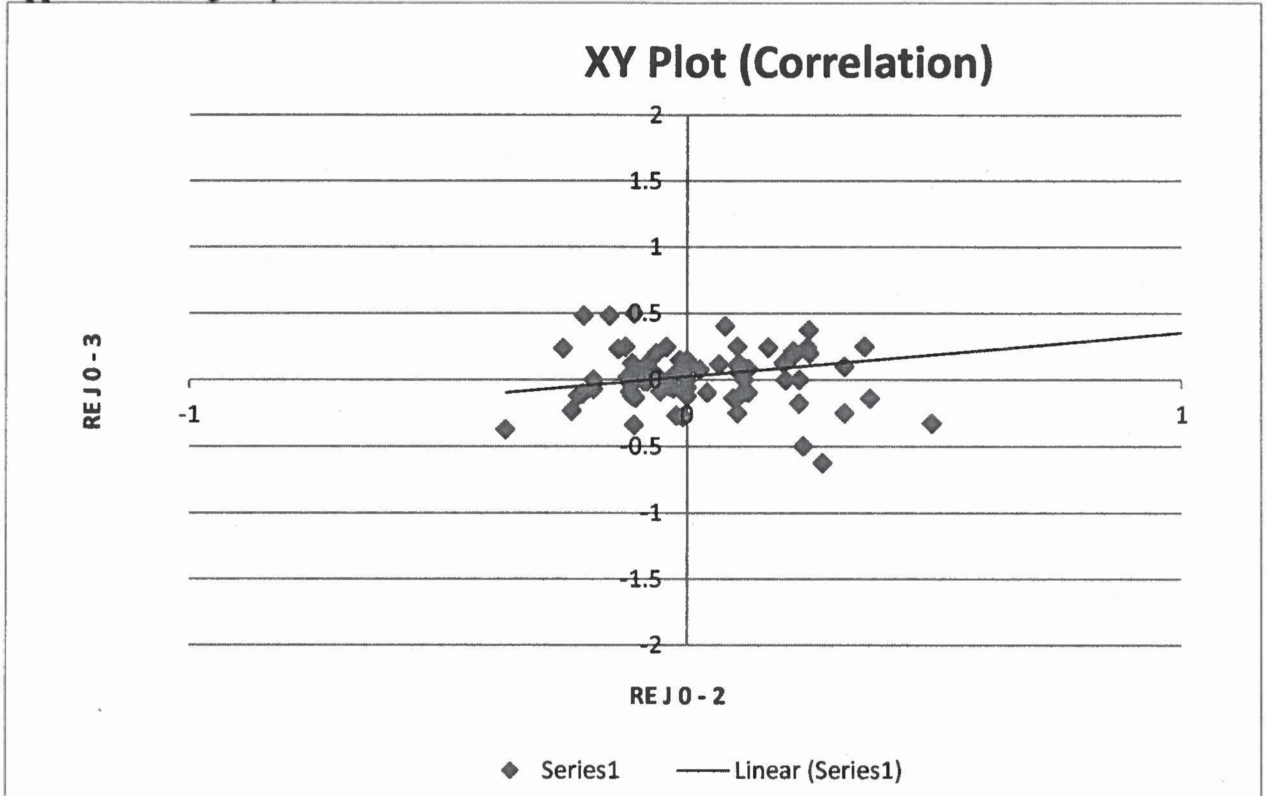
Appendix B

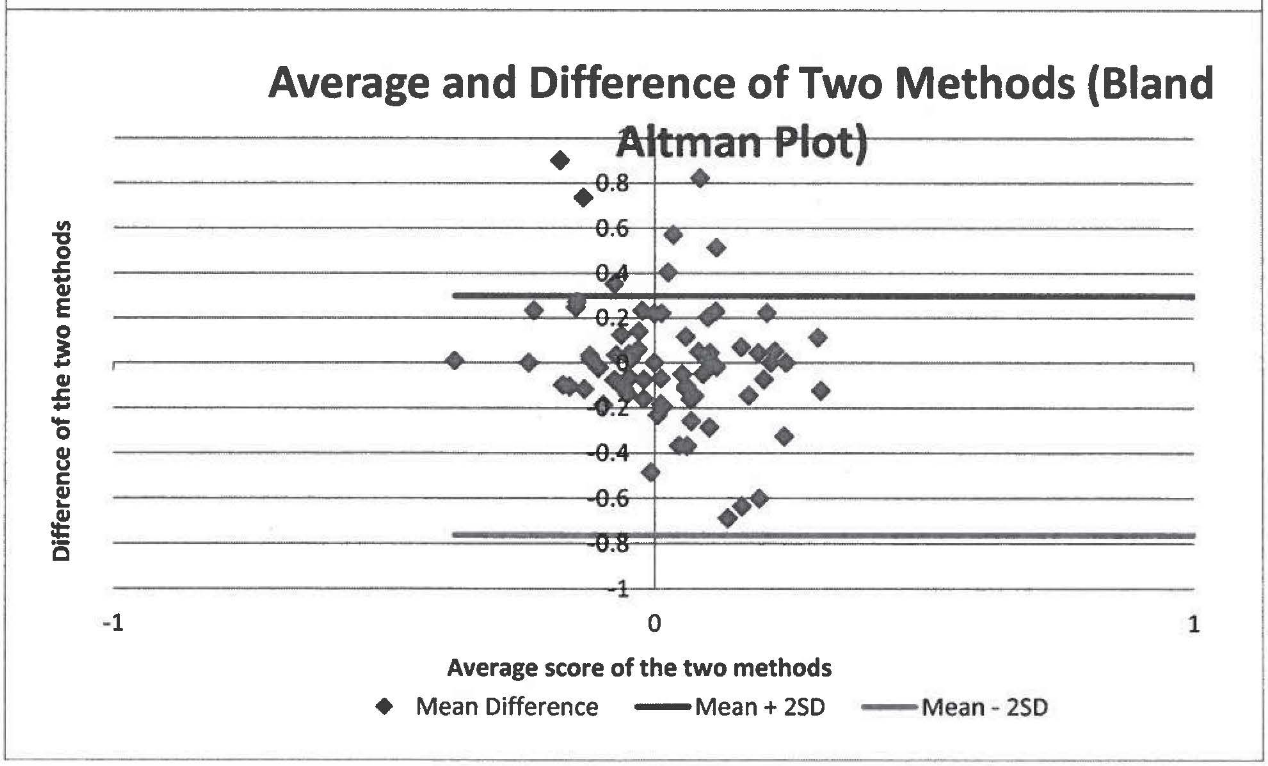
Appendix B: Right eye, J0, instrument 1 versus instrument 2



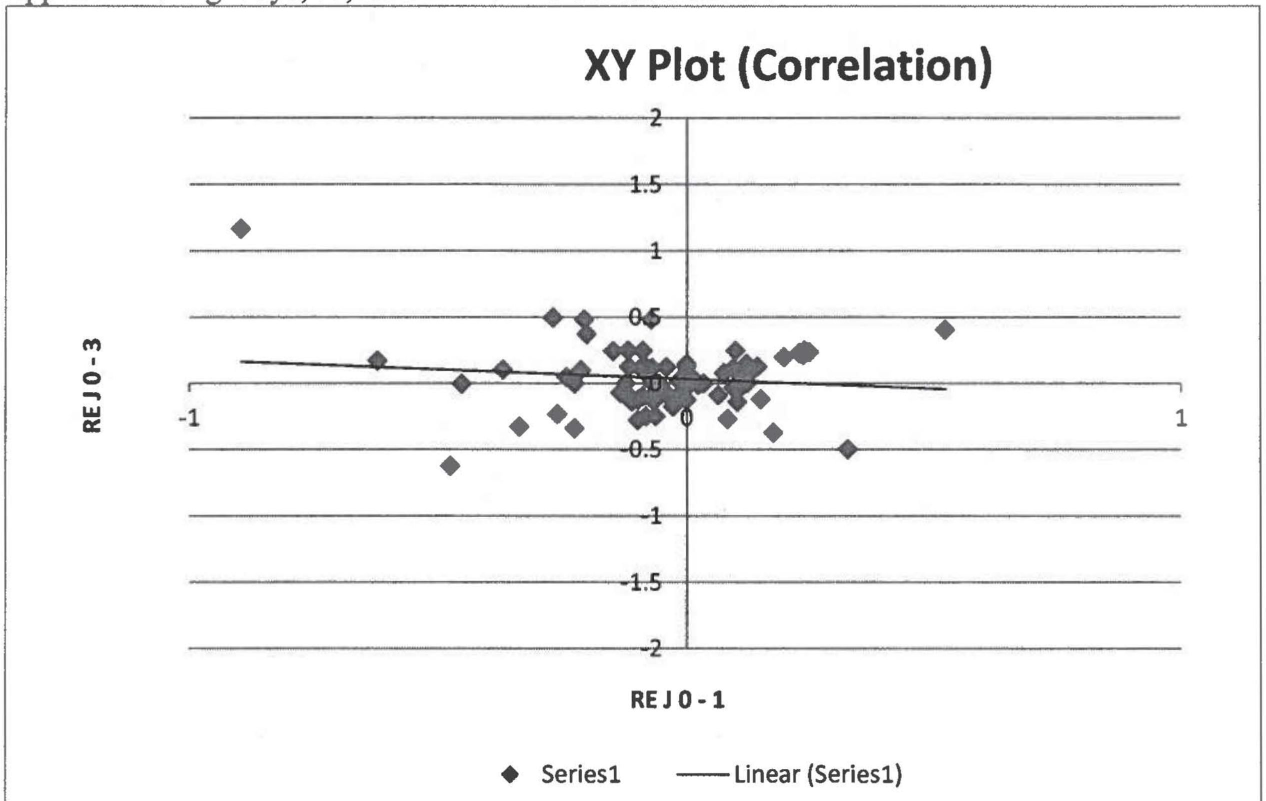


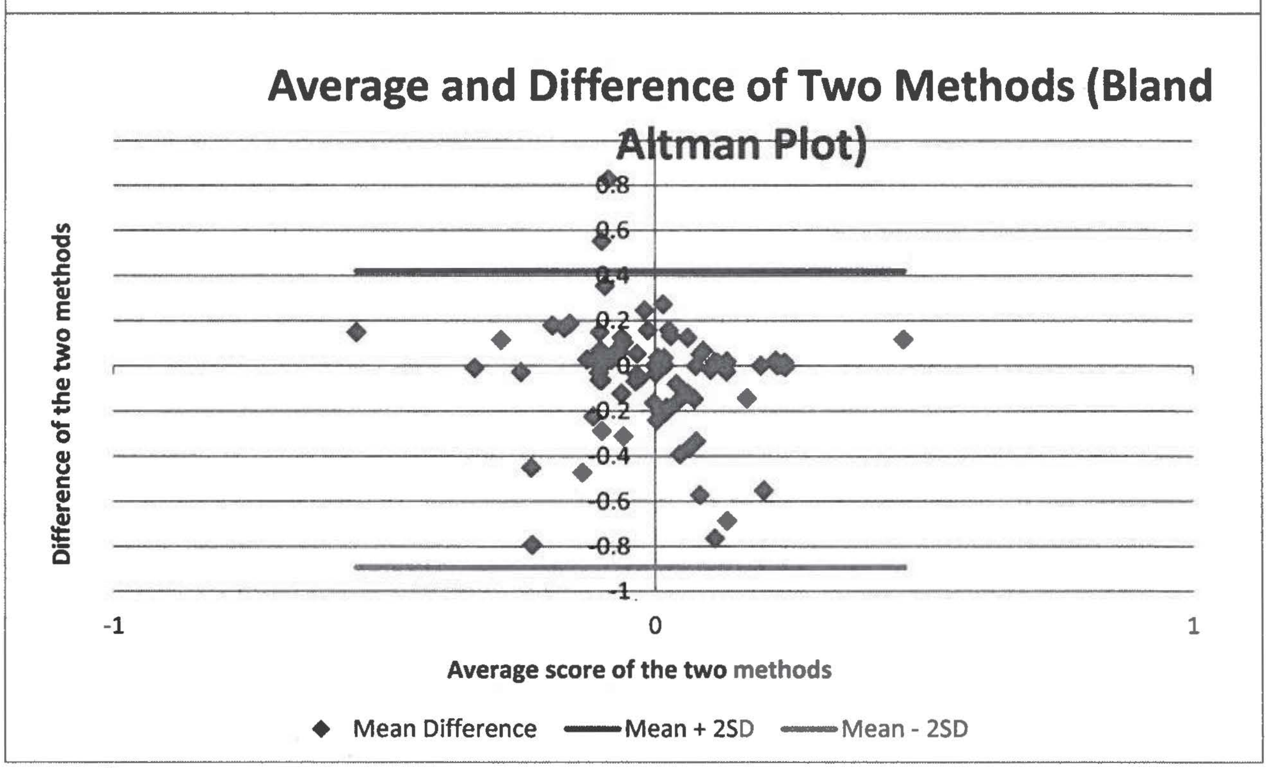
Appendix B: Right eye, J0, instrument 2 versus instrument 3



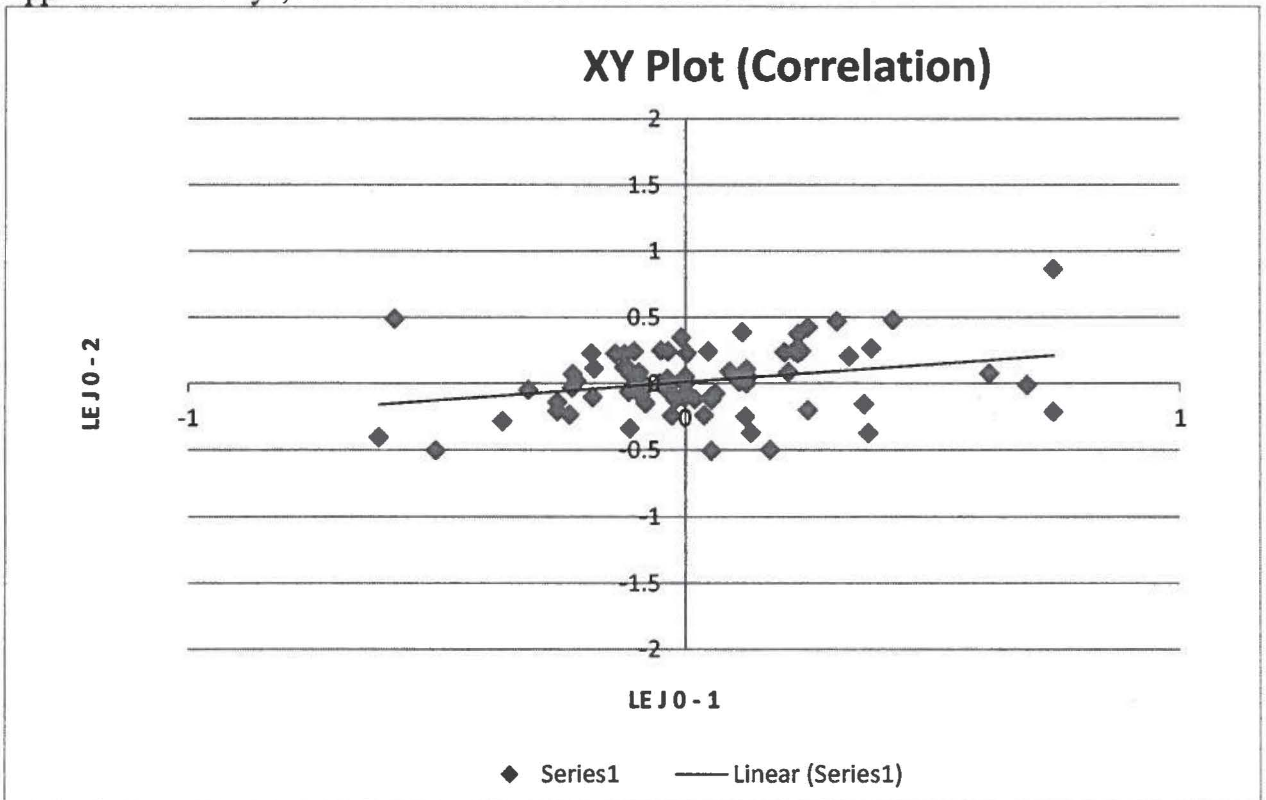


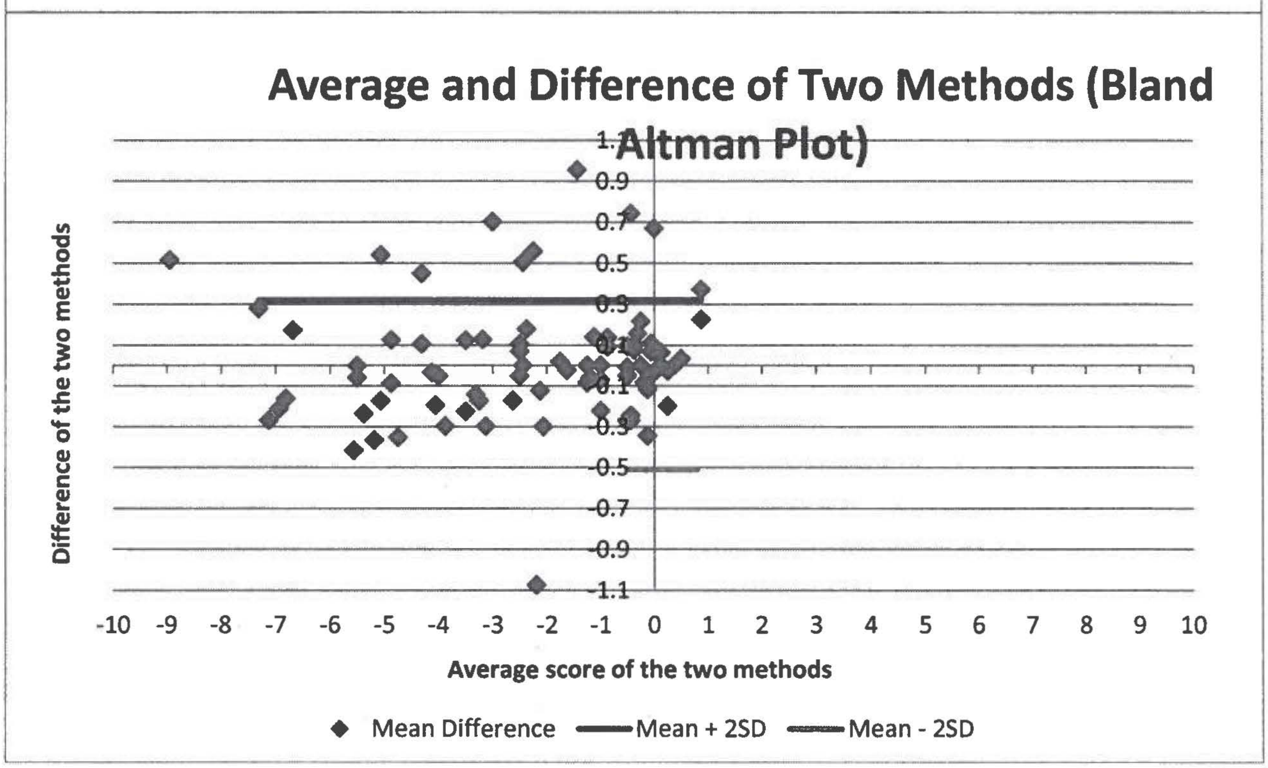
Appendix B: Right eye, J0, instrument 1 versus instrument 3



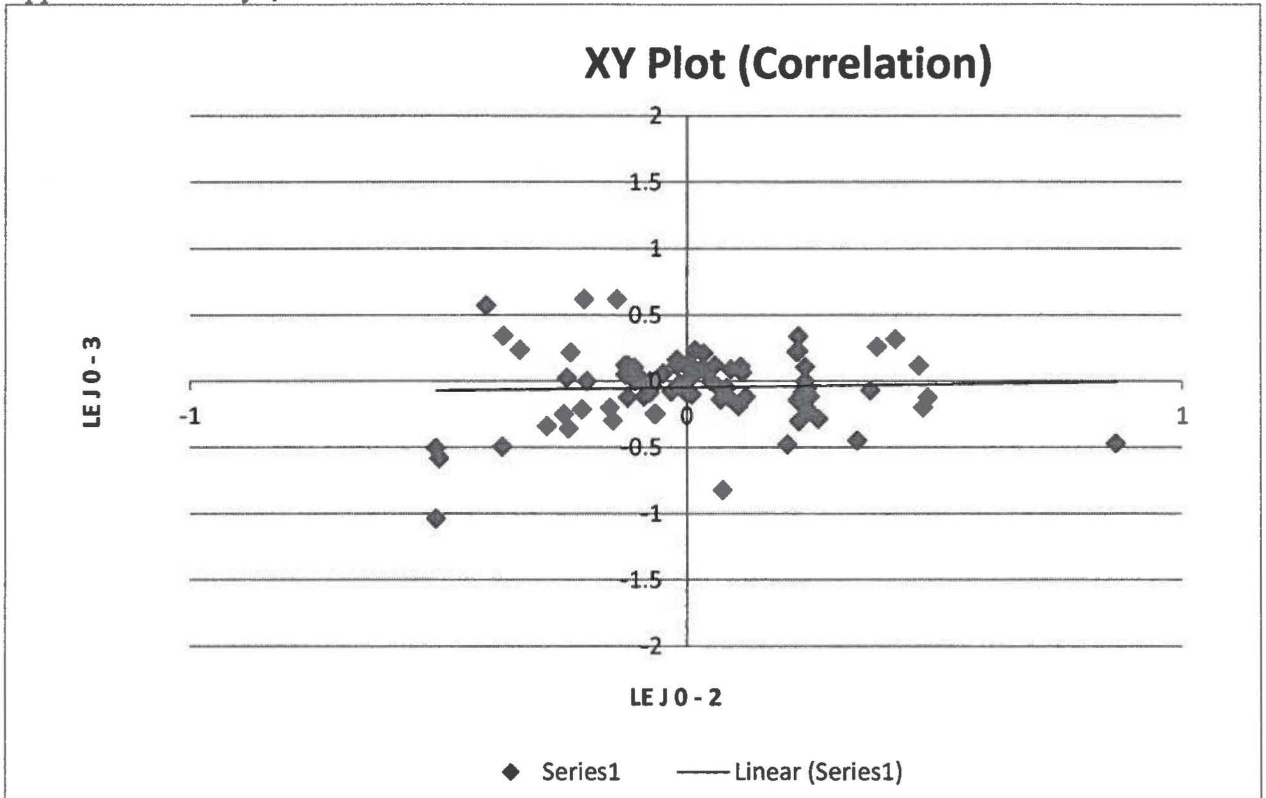


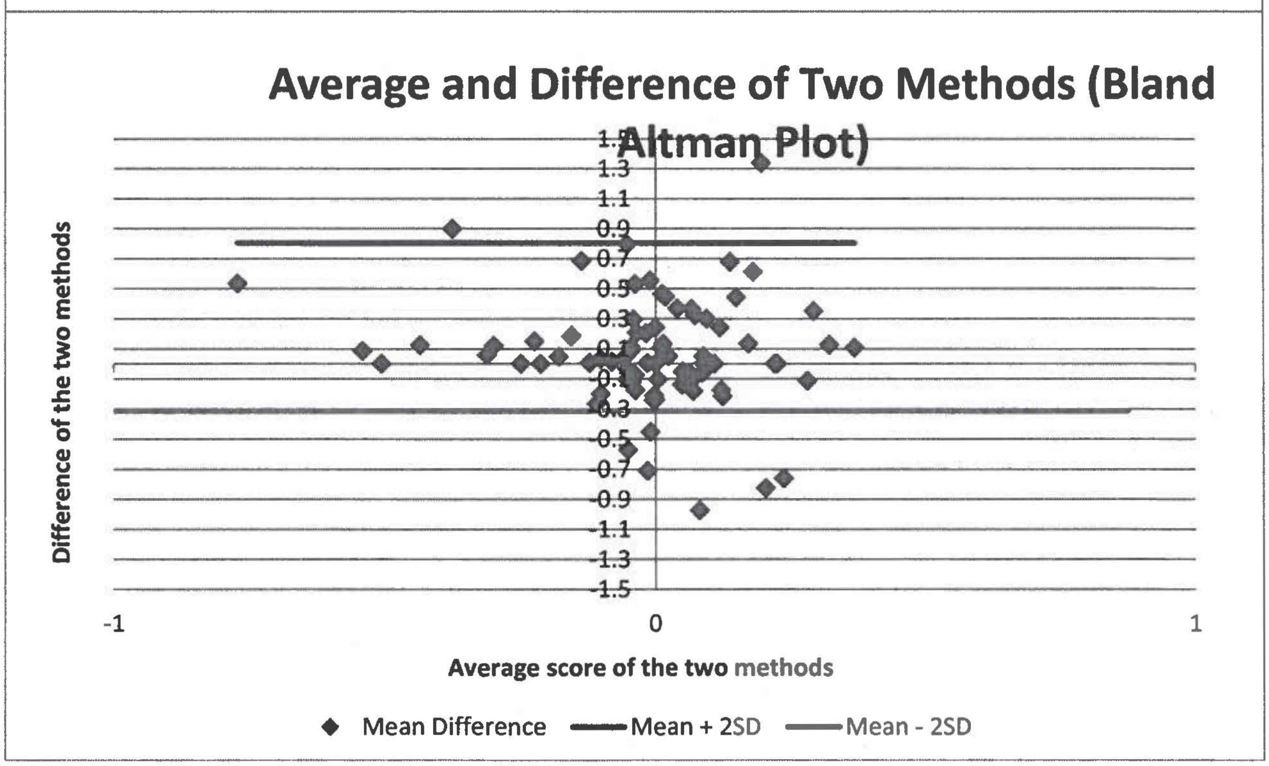
Appendix B: Left eye, J0 instrument 1 versus instrument 2



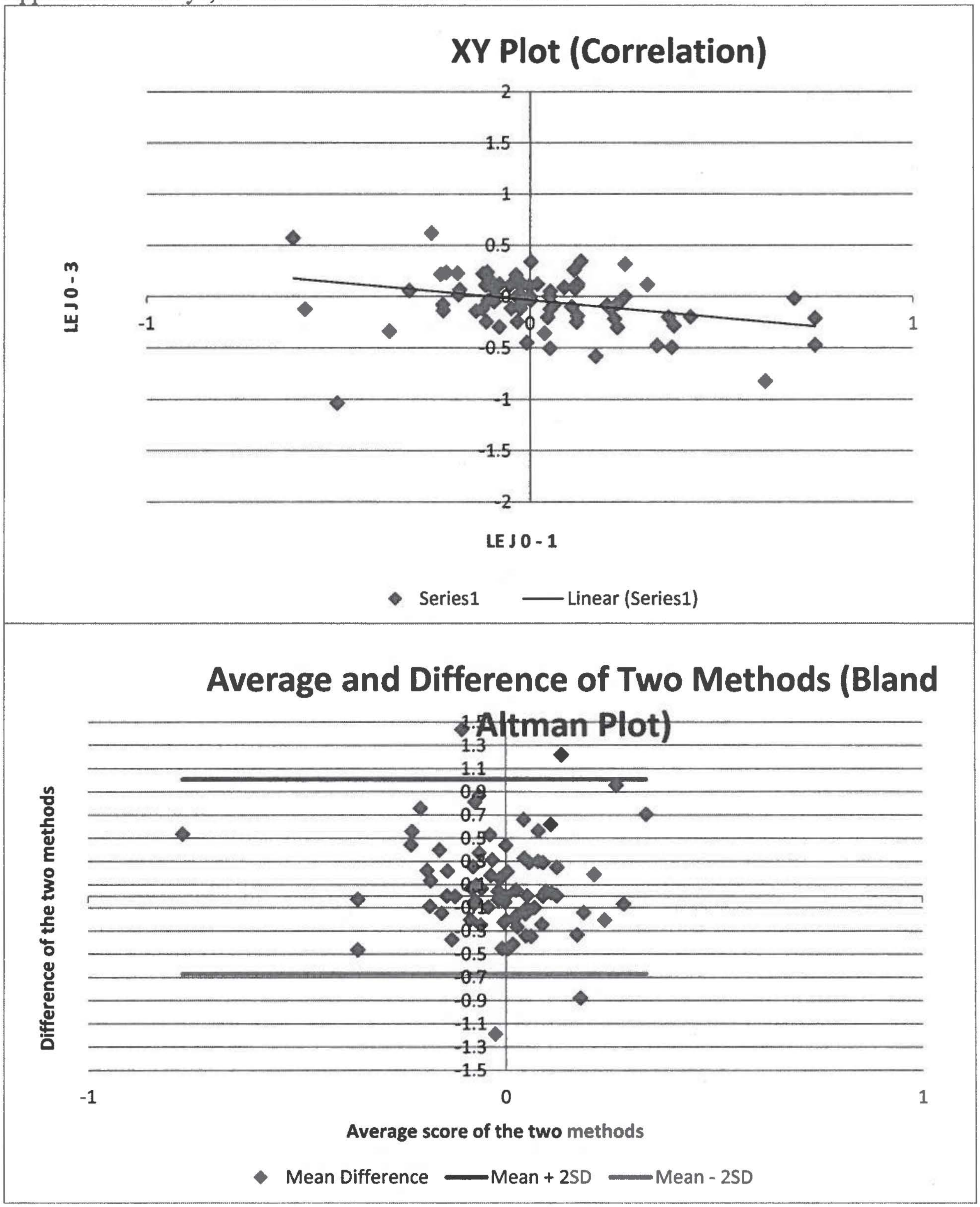


Appendix B: Let eye, J0 instrument 2 versus instrument 3



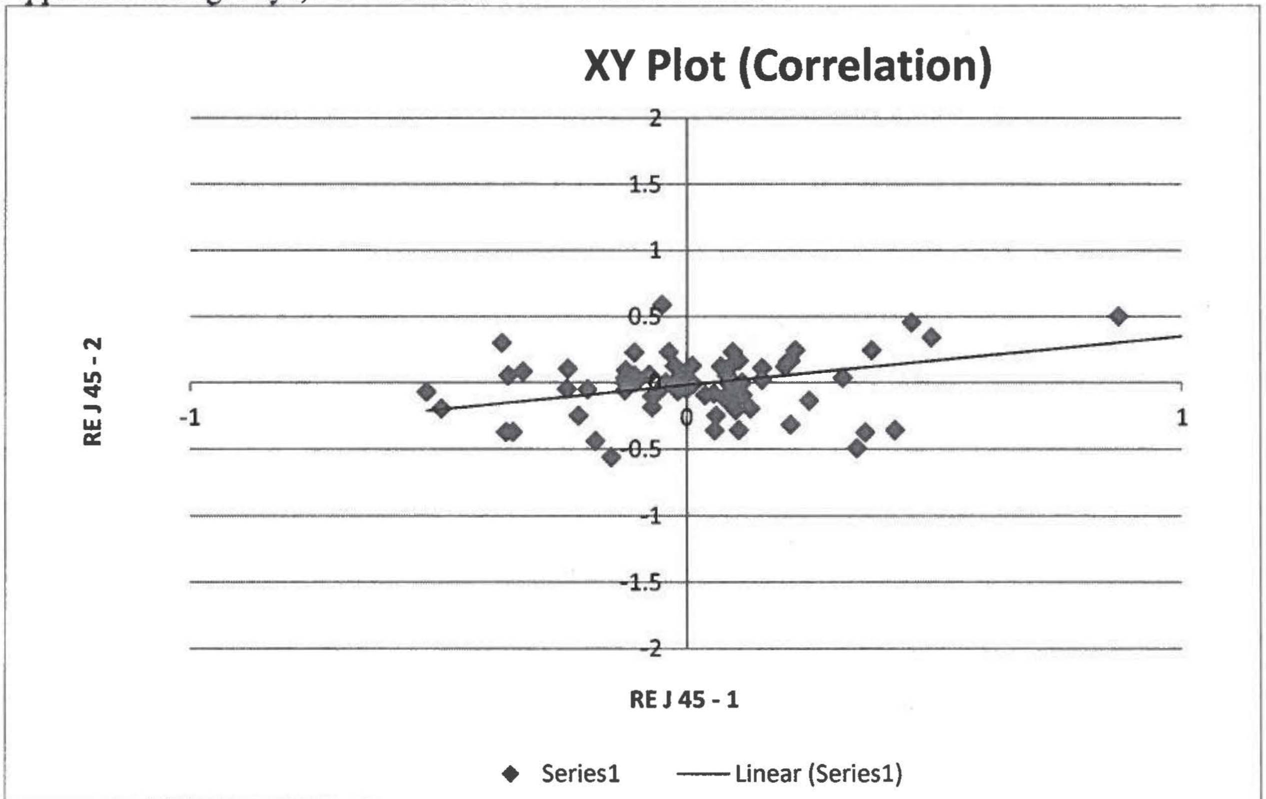


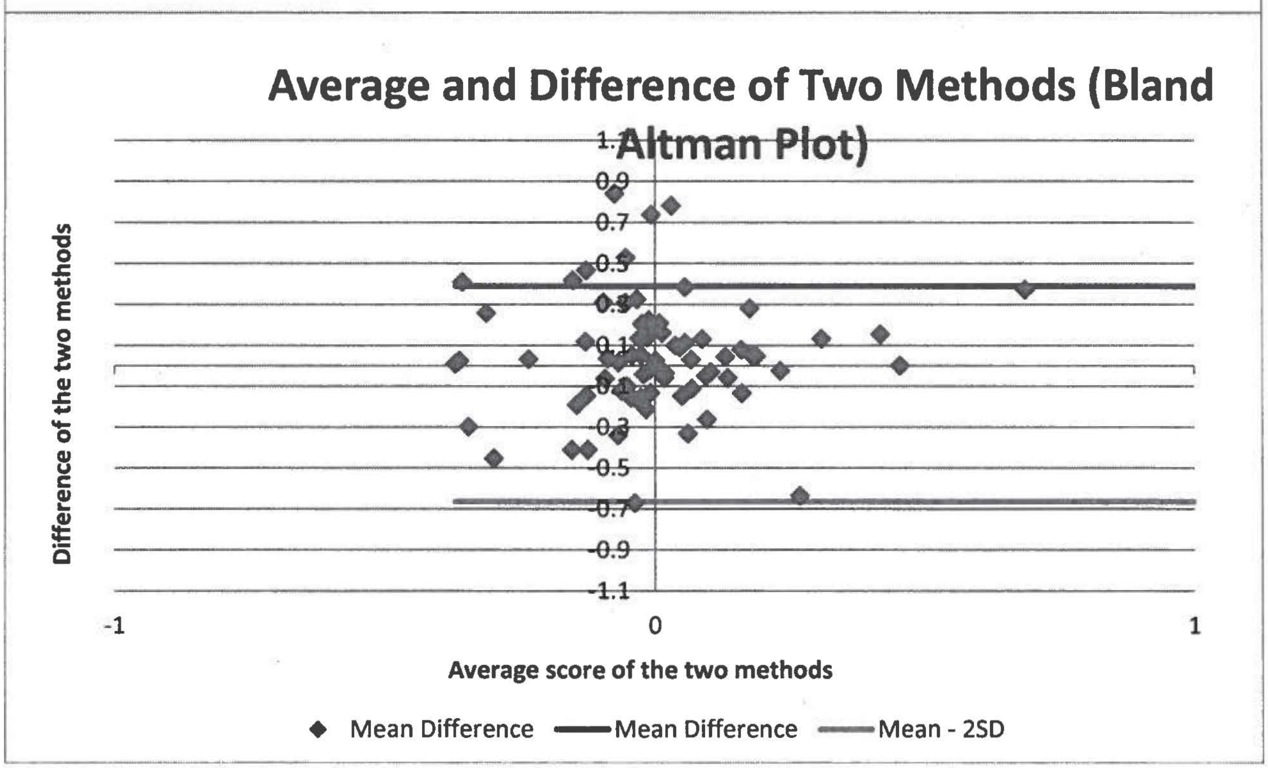
Appendix B: Let eye, J0 instrument 1 versus instrument 3



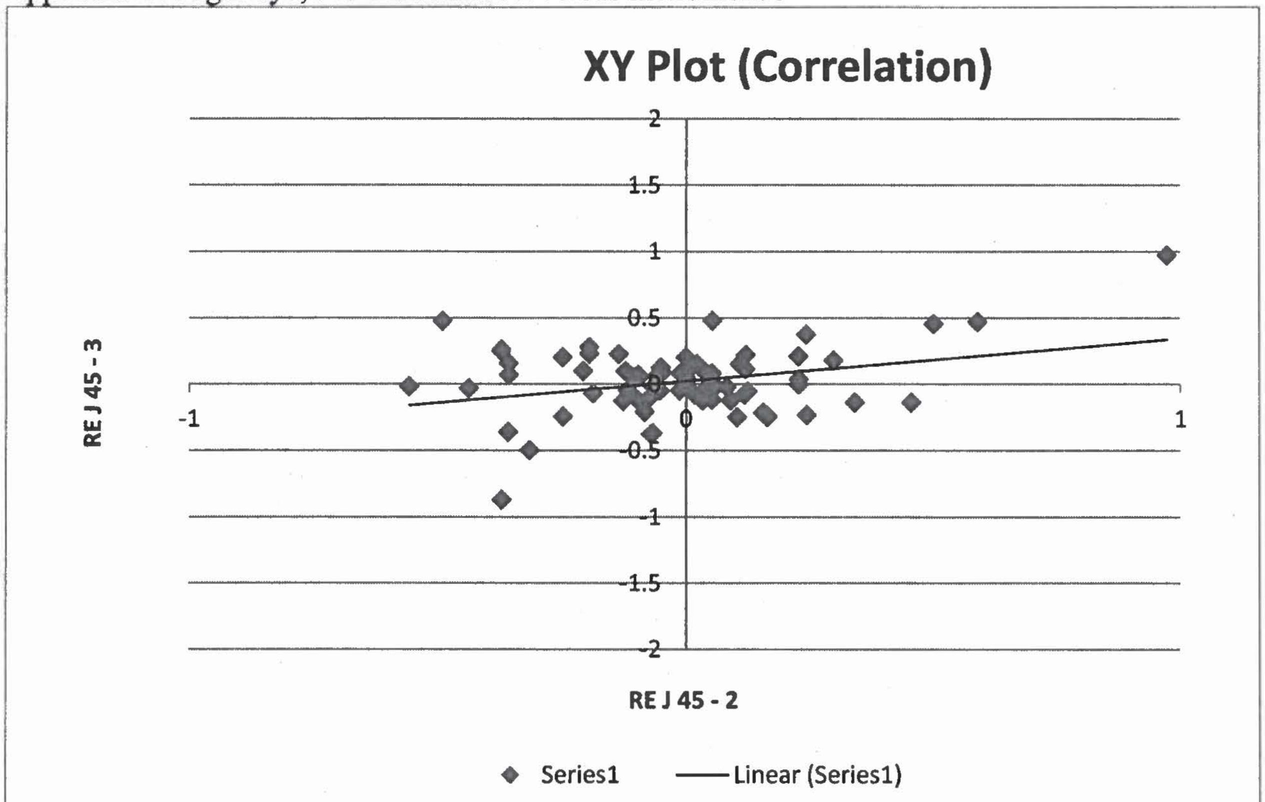
Appendix C

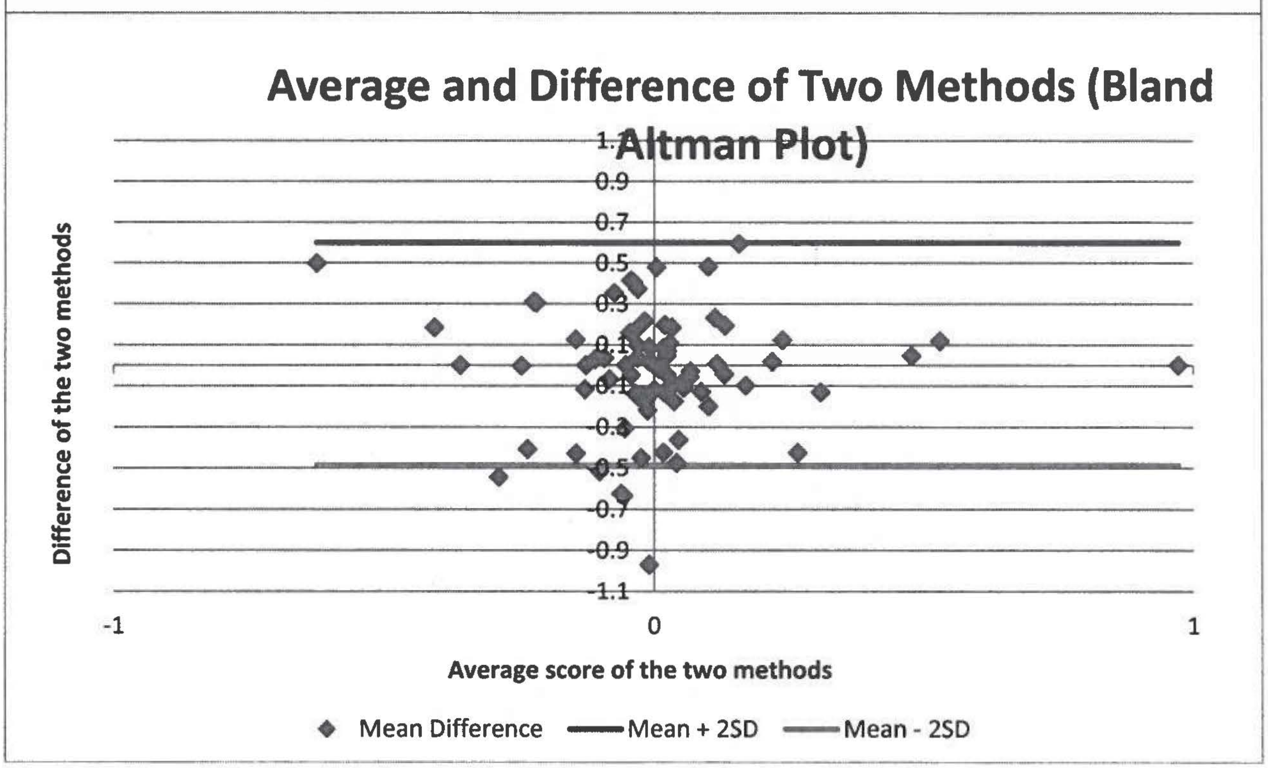
Appendix C: Right eye, J45 instrument 1 versus instrument 2



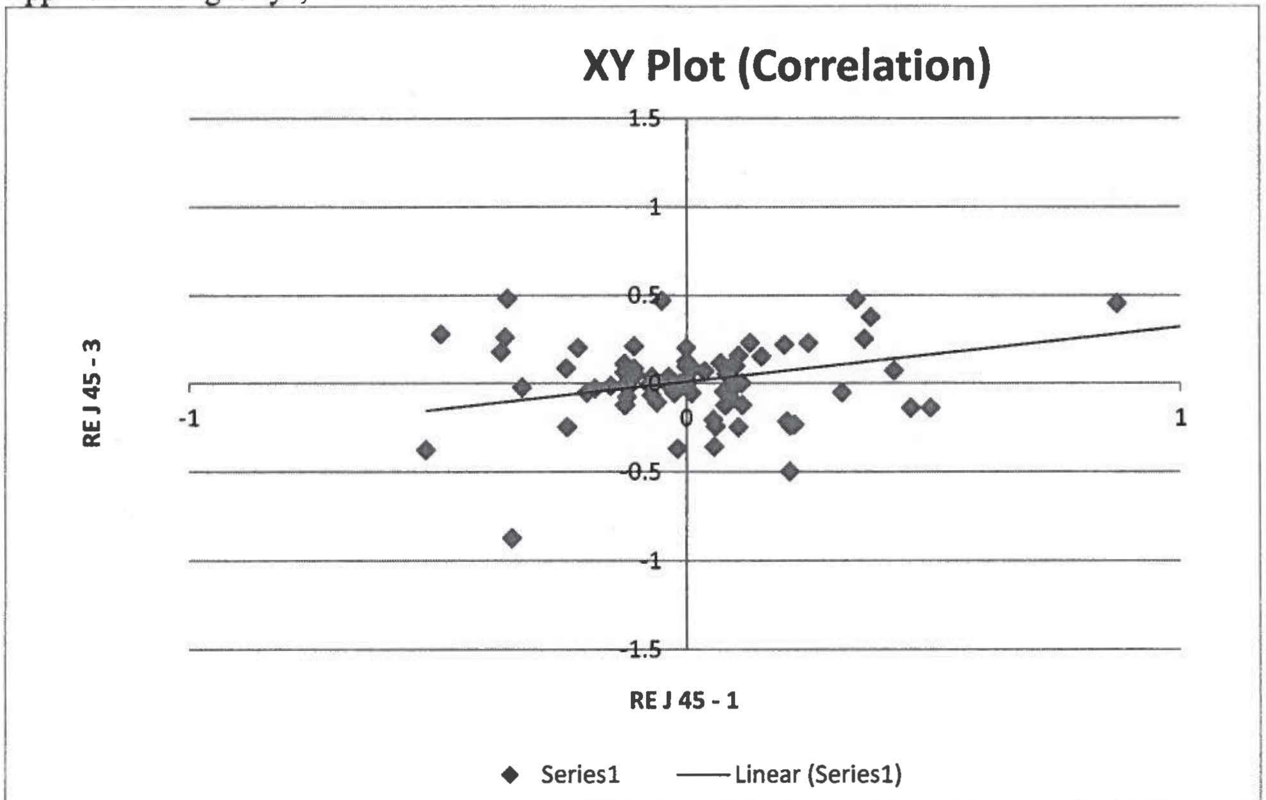


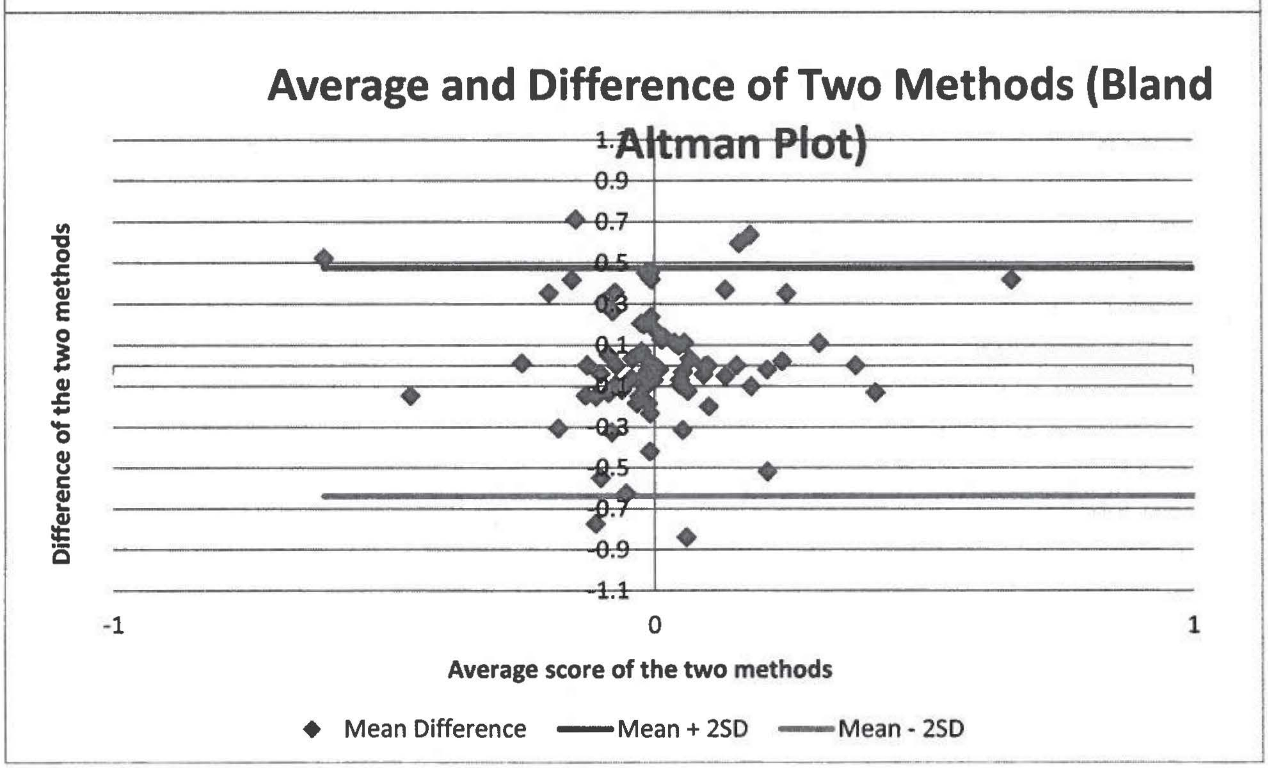
Appendix C: Right eye, J45 instrument 2 versus instrument 3



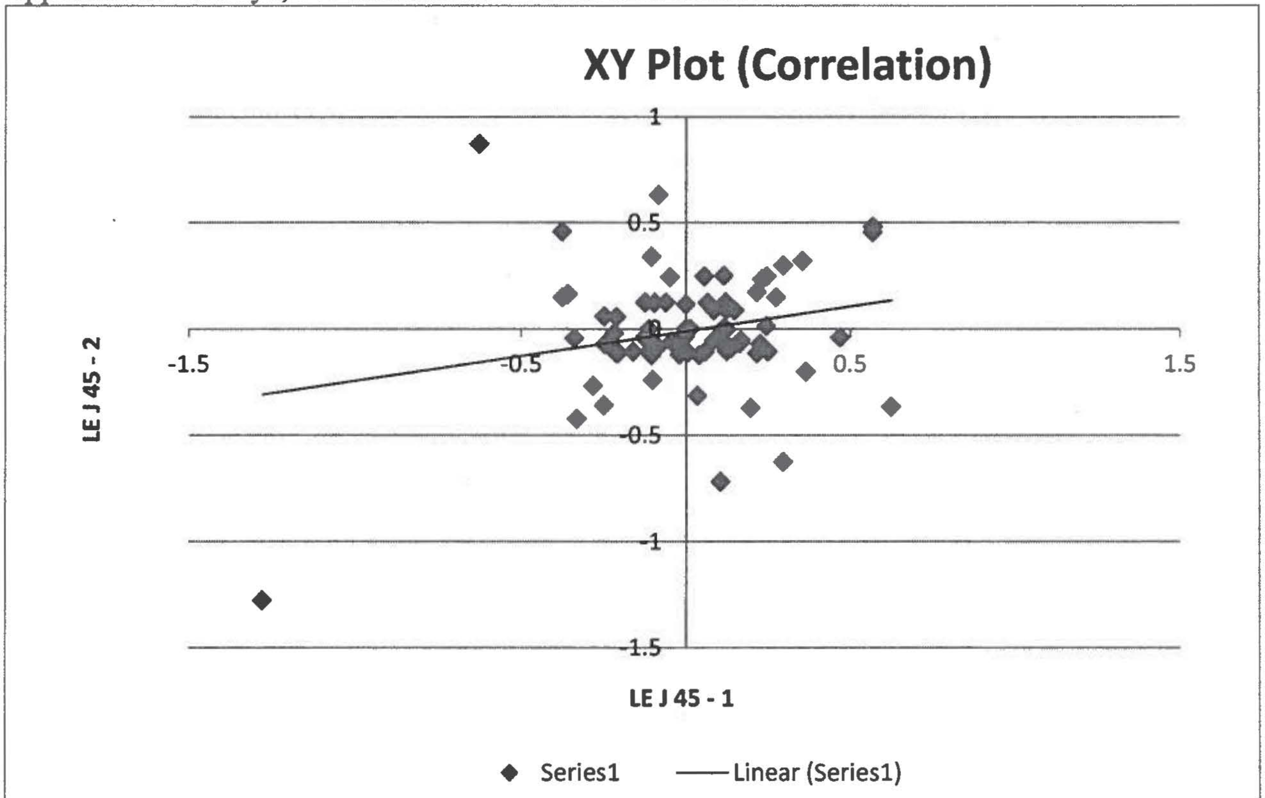


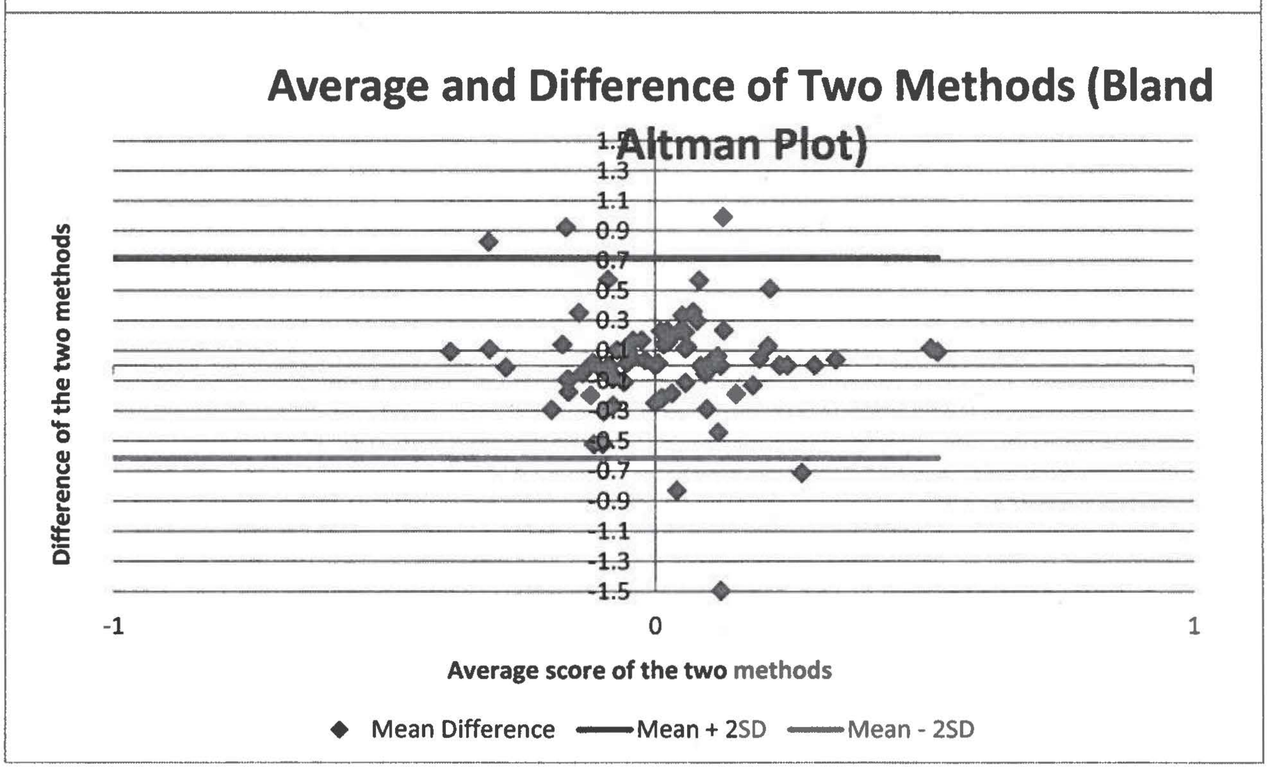
Appendix C: Right eye, J45 instrument 1 versus instrument 3



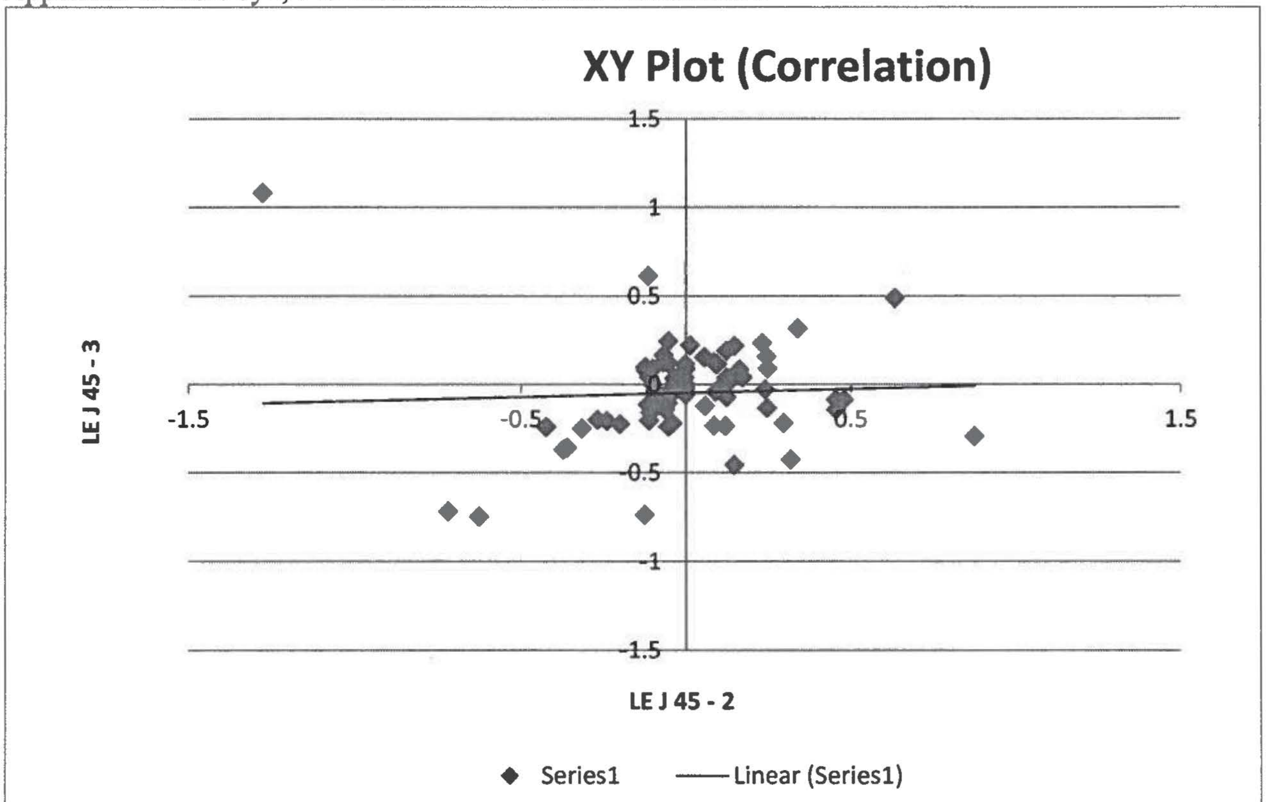


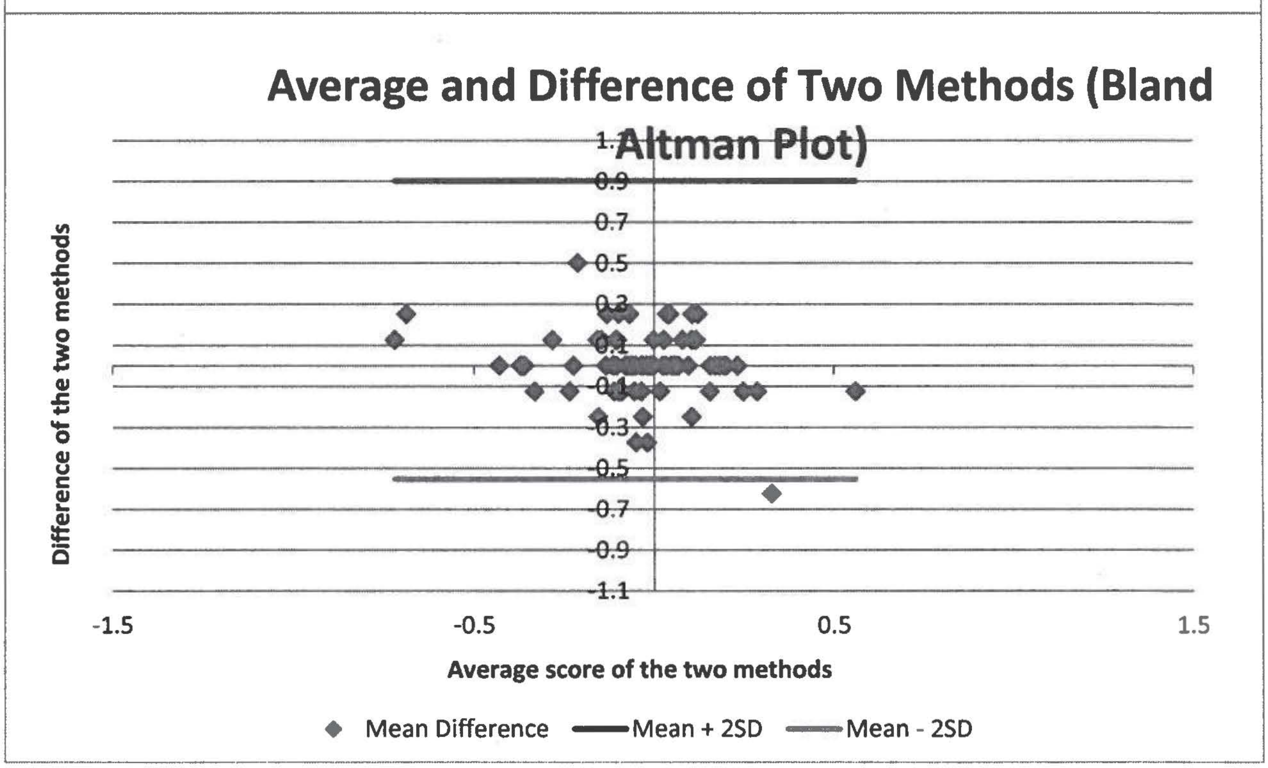
Appendix C: Left eye, J45 instrument 1 versus instrument 2



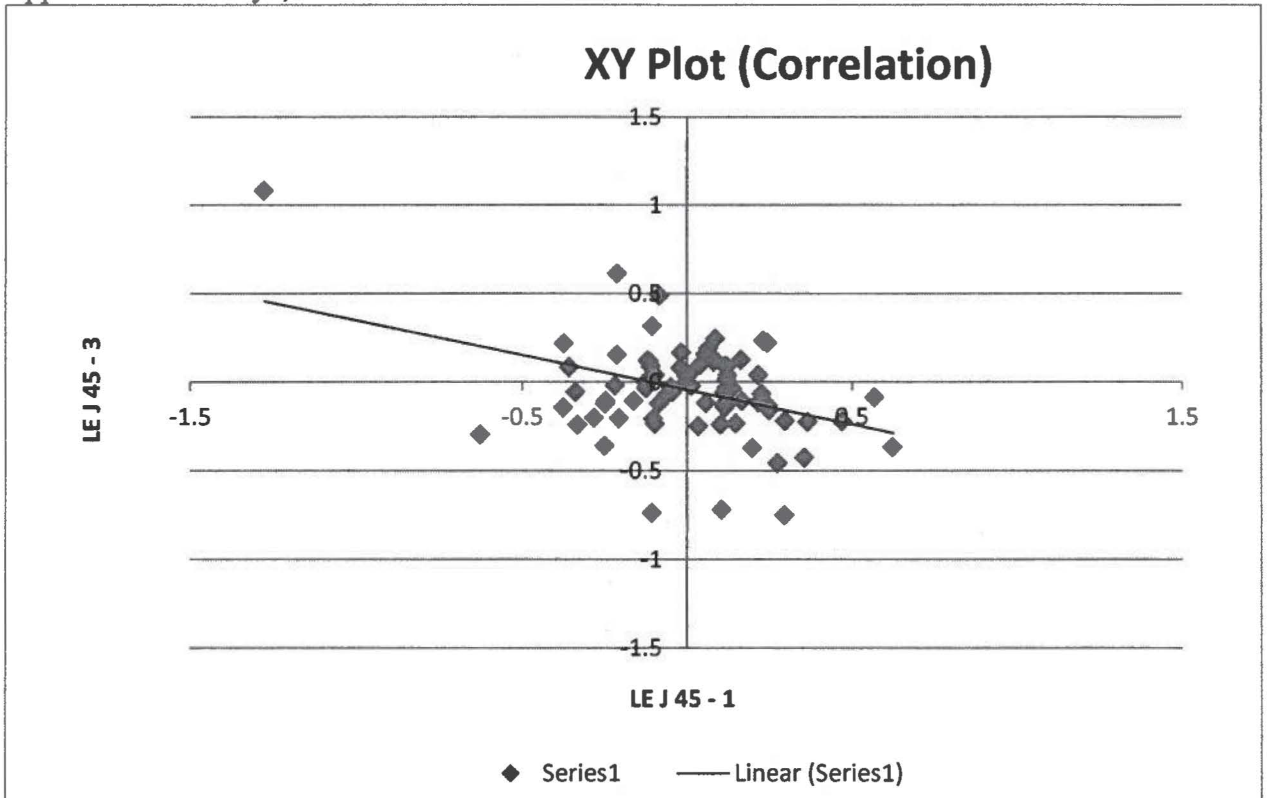


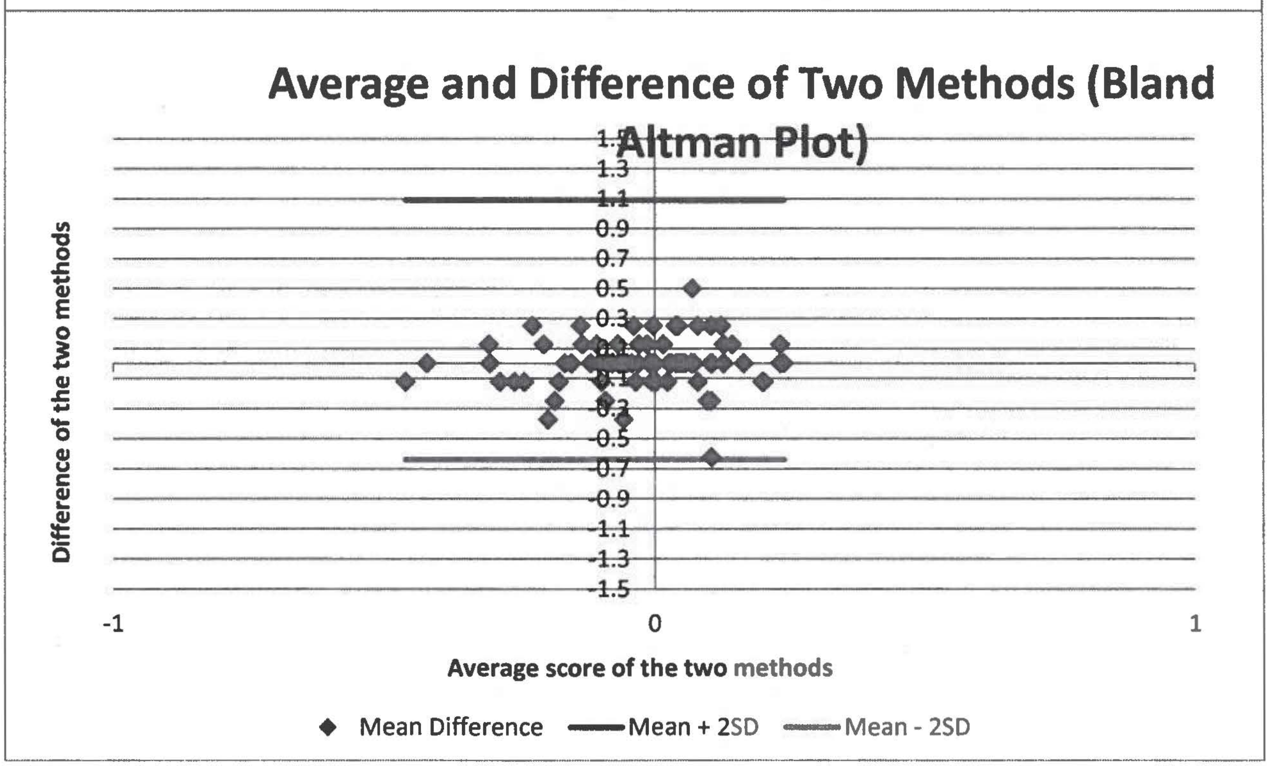
Appendix C: Left eye, J45 instrument 2 versus instrument 3





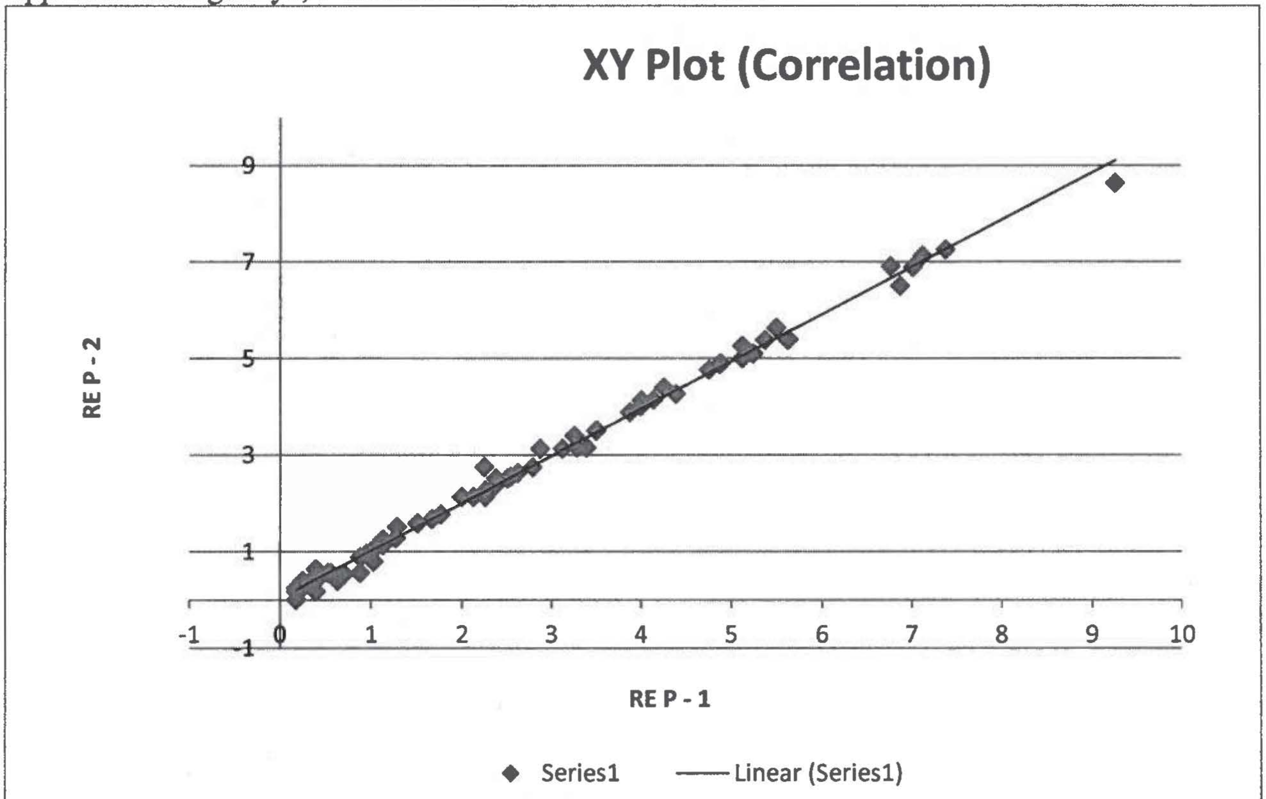
Appendix C: Left eye, J45 instrument 1 versus instrument 3

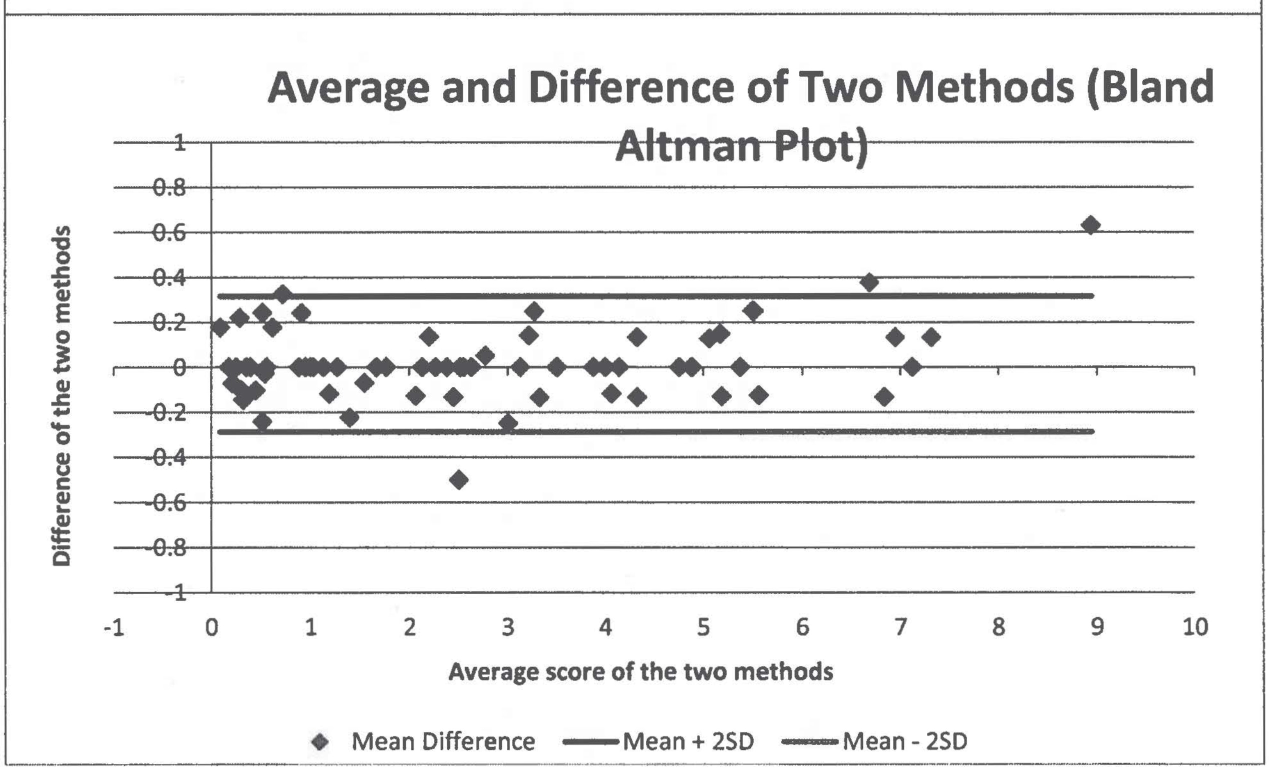




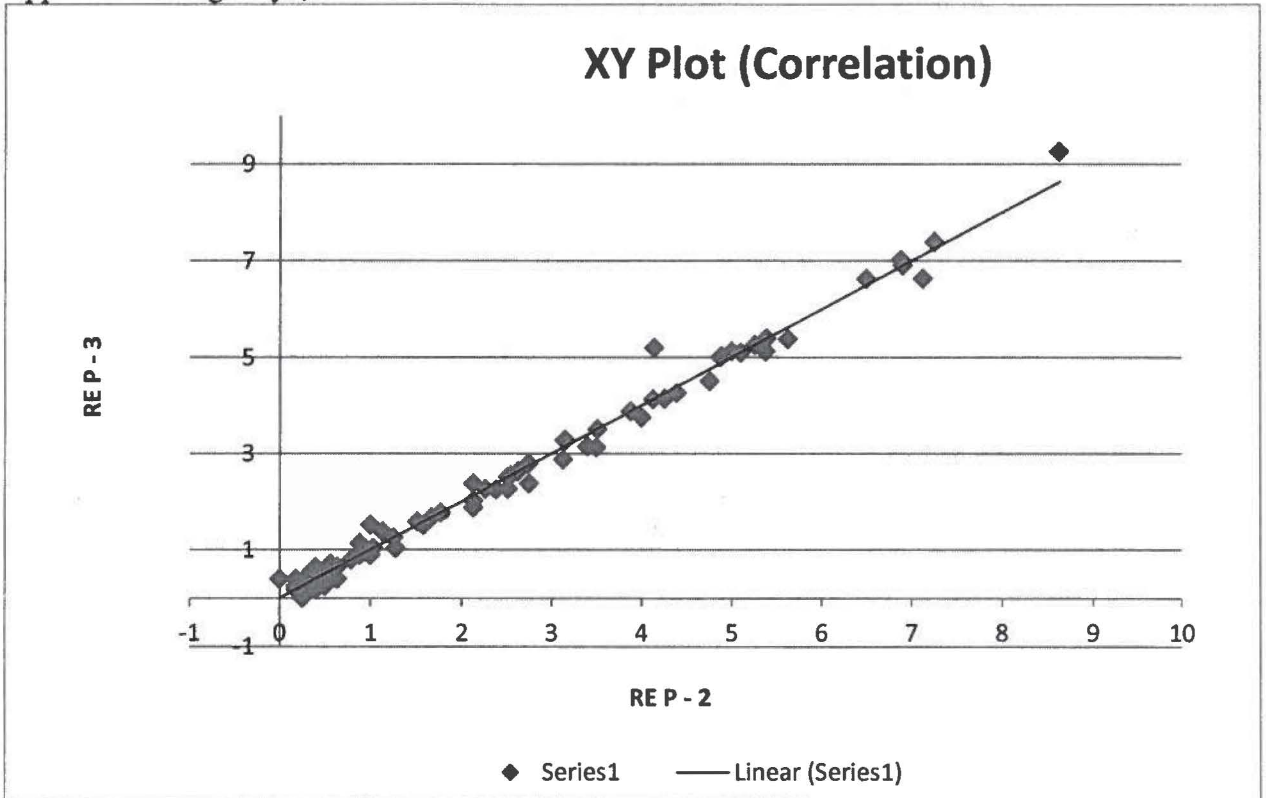
Appendix D

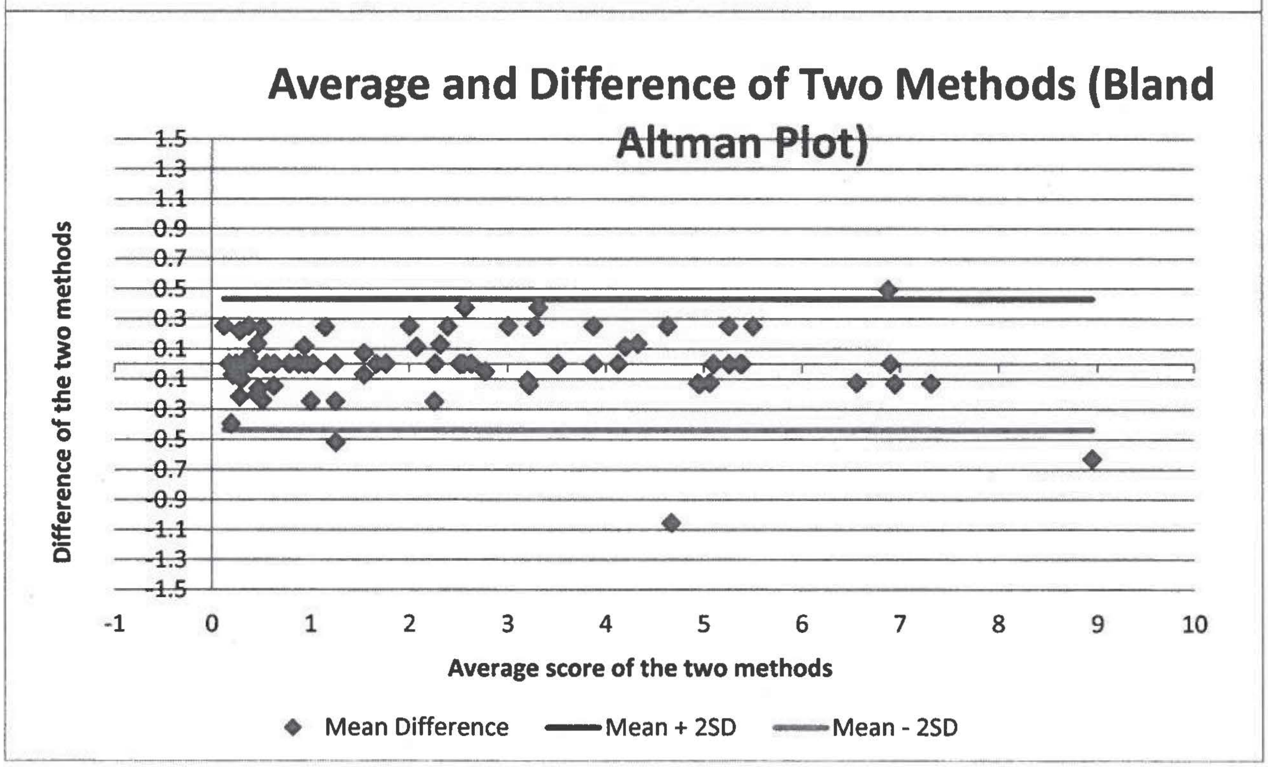
Appendix D: Right eye, P instrument 1 versus instrument 2



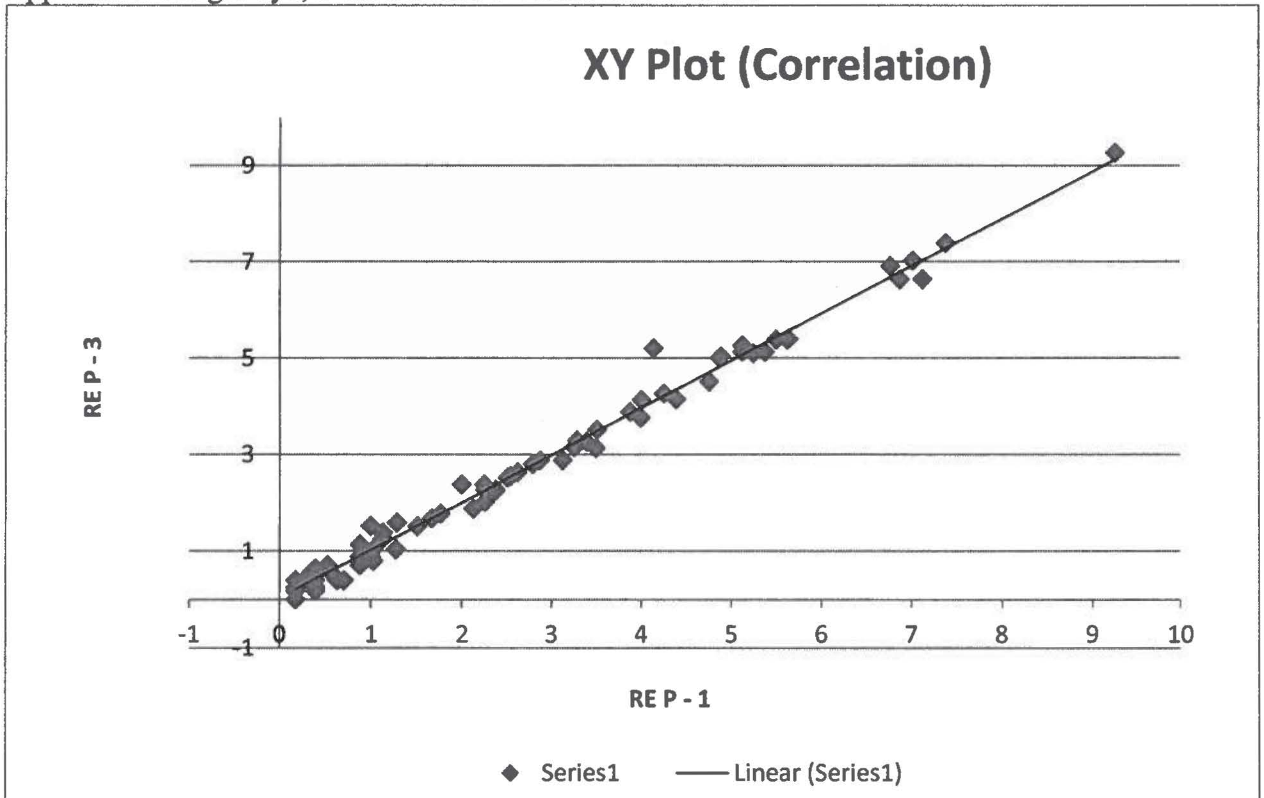


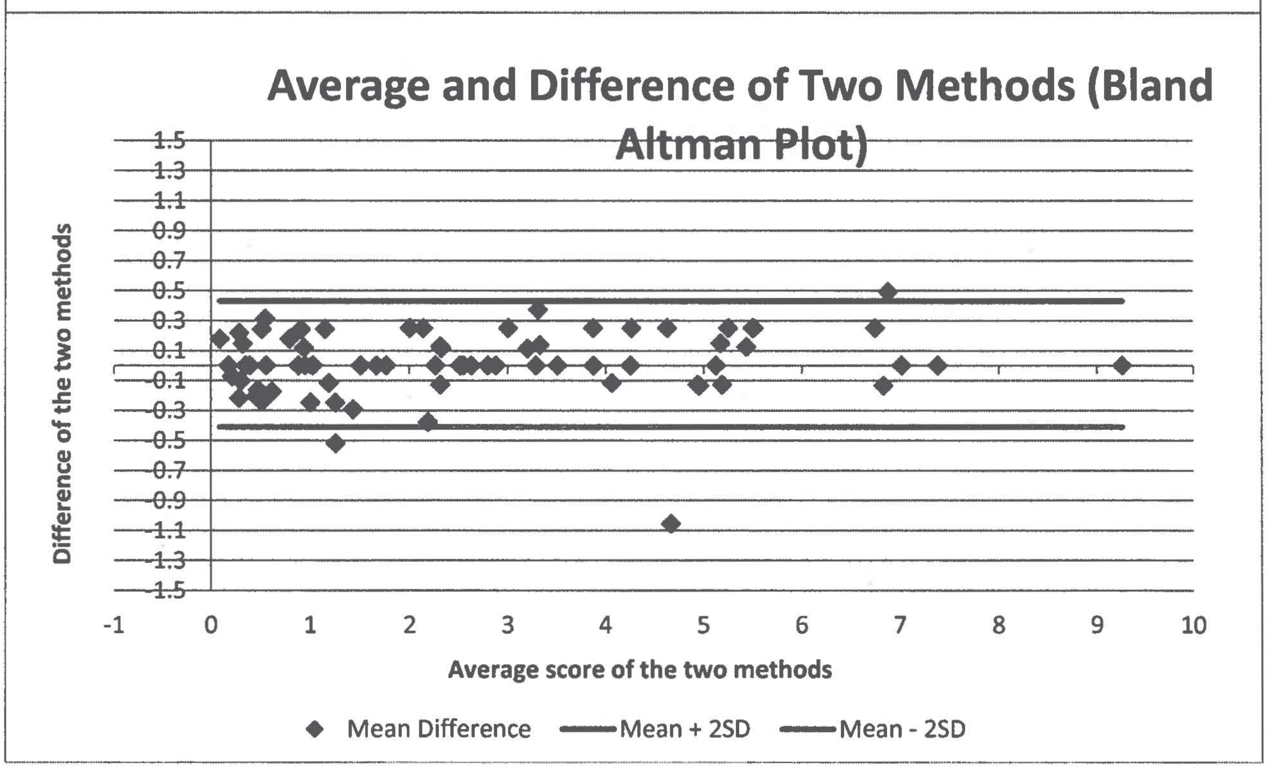
Appendix D: Right eye, P instrument 2 versus instrument 3



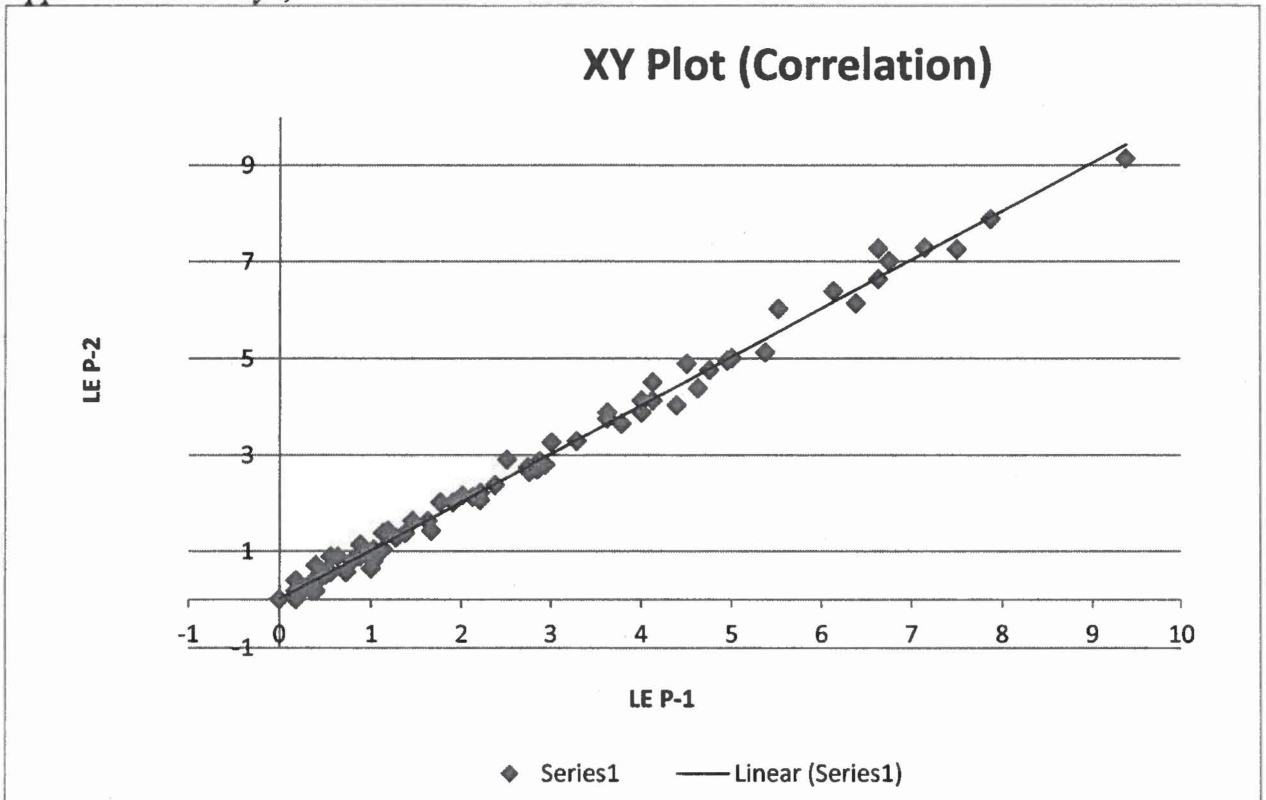


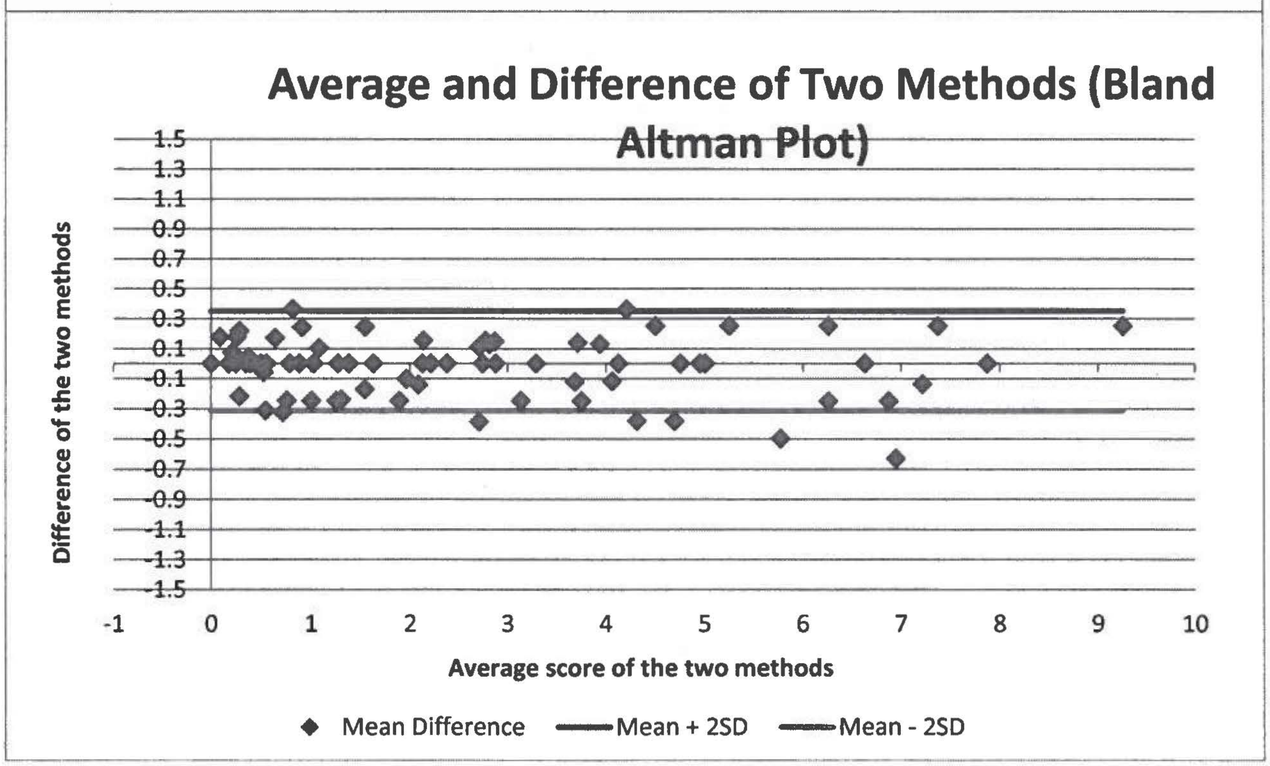
Appendix D: Right eye, P instrument 1 versus instrument 3



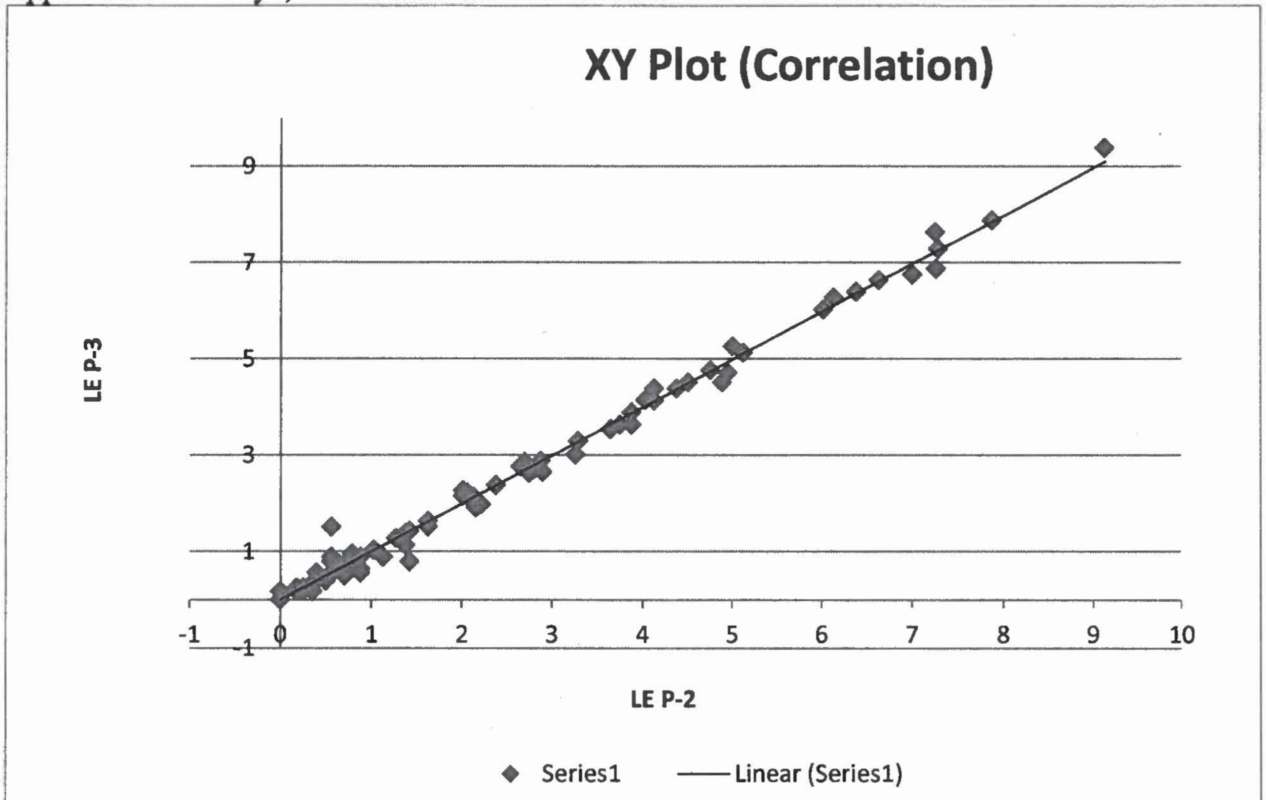


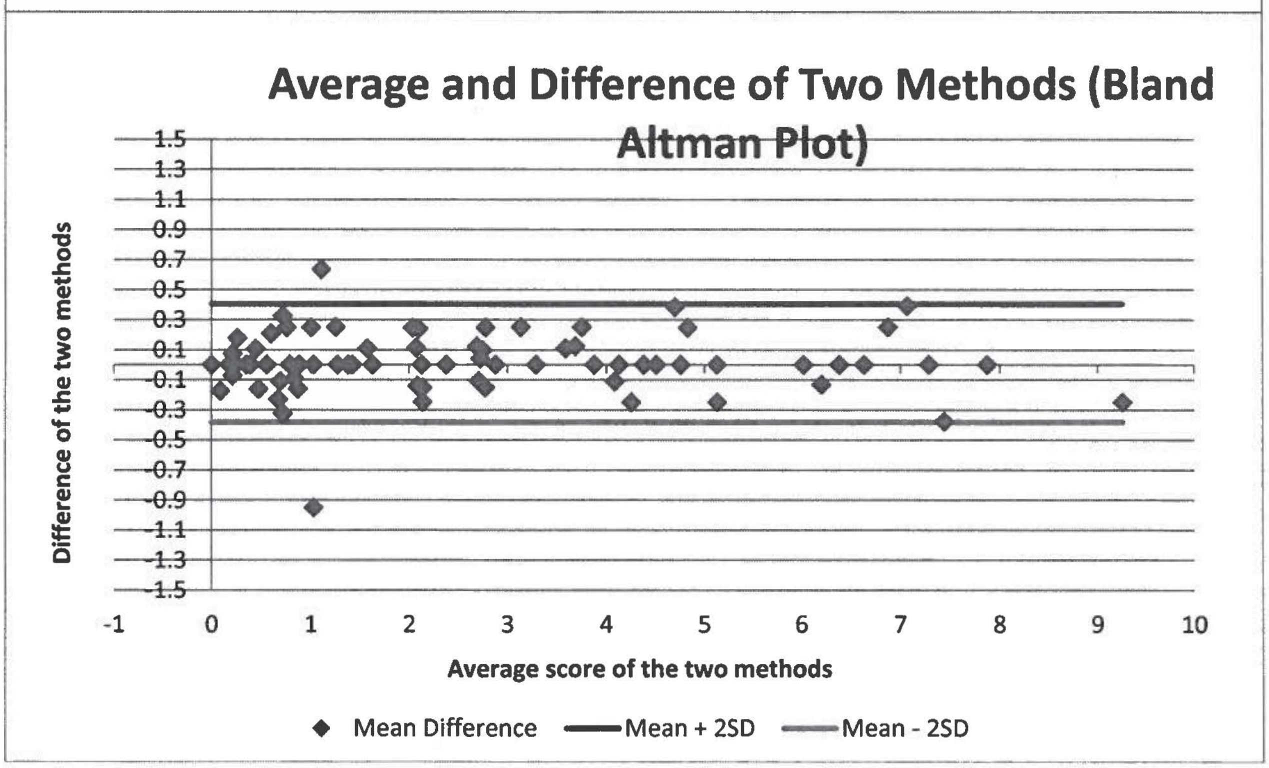
Appendix D: Left eye, P instrument 1 versus instrument 2



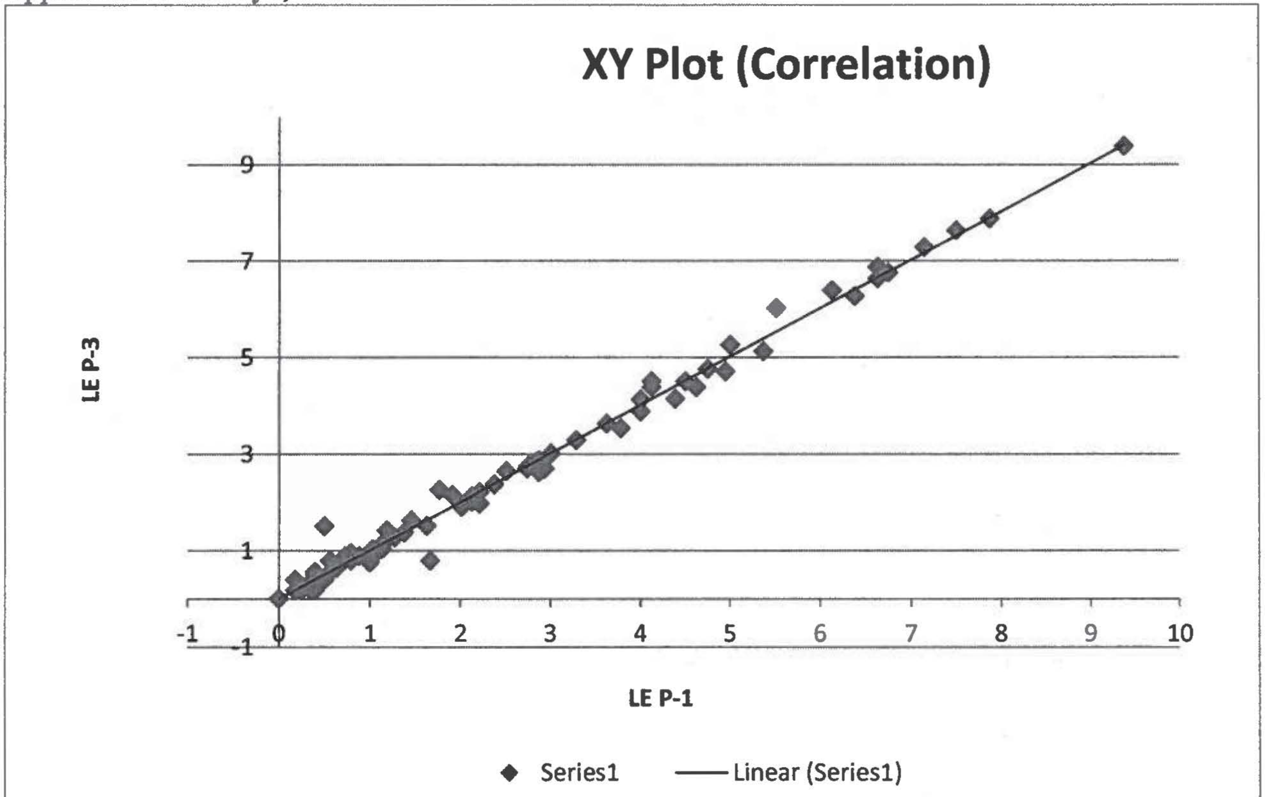


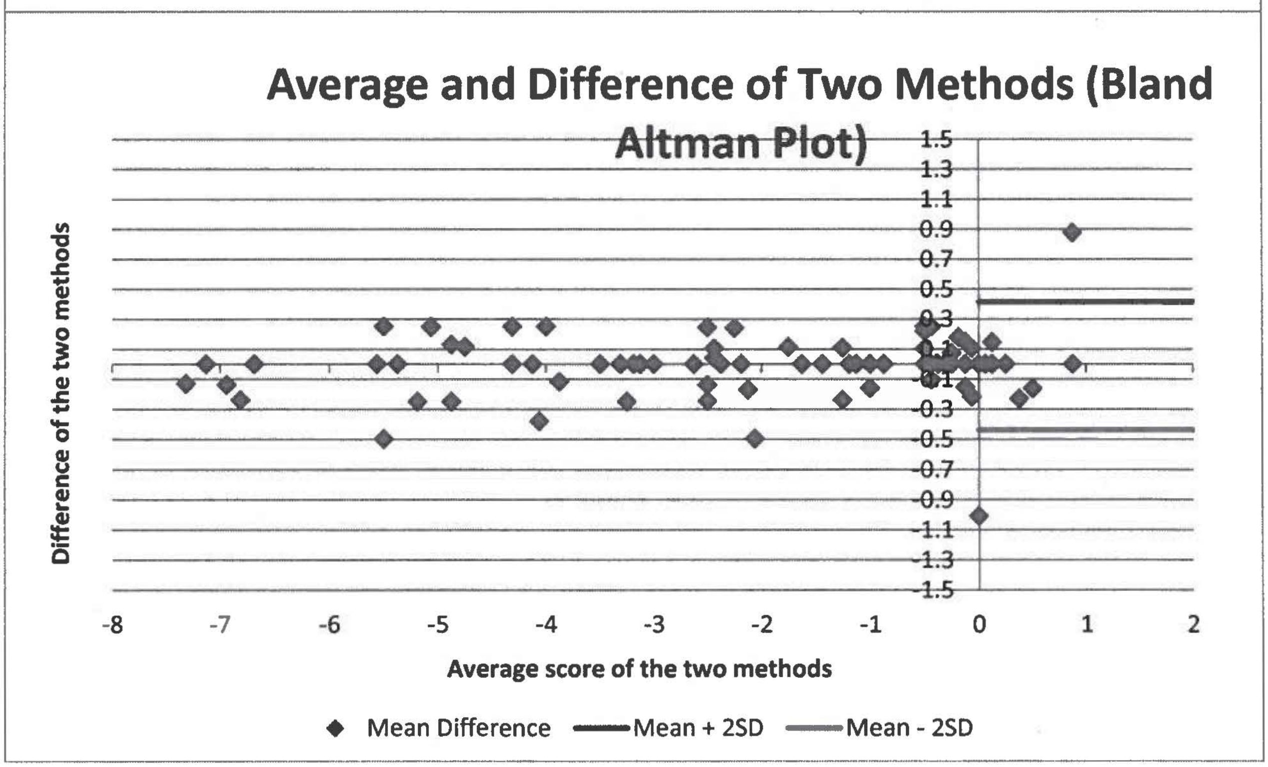
Appendix D: Left eye, P instrument 2 versus instrument 3





Appendix D: Left eye, P instrument 1 versus instrument 3





Appendix E

Appendix E: Mean, Standard Deviation, and 1.96 Standard Deviation Summary Tables

	RE M-1 : M-2	RE M-2 : M-3	RE M-1 : M-3	LE M-1 : M-2	LE M-2 : M-3	LE M-1 : M-3
Mean	0.00304878	-0.028963415	-0.025914634	0.056402439	-0.021341463	0.035060976
St Dev	0.161941835	0.226775216	0.224586215	0.212597508	0.238795911	0.240782538
1.96 St Dev	0.317405998	0.444479424	0.440188981	0.416691115	0.468039985	0.471933775

	RE J0-1 : J0-2	RE J0-2 : J0-3	RE J0-1 : J0-3	LE J0-1 : J0-2	LE J0-2 : J0-3	LE J0-1 : J0-3
Mean	-0.073024543	0.000339321	-0.072685222	0.010542353	0.064459998	0.075002351
St Dev	0.350429969	0.27093402	0.335267789	0.297282893	0.361412846	0.42813457
1.96 St Dev	0.686842739	0.531030679	0.657124867	0.58267447	0.708369178	0.839143757

	RE J45-1 : J45-2	RE J45-2 : J45-3	RE J45-1 : J45-3	LE J45-1 : J45-2	LE J45-2 : J45-3	LE J45-1 : J45-3
Mean	0.03878931	-0.026804518	0.011984791	0.020072831	0.041868386	0.061941216
St Dev	0.268943116	0.277337404	0.283903237	0.340171026	0.371485292	0.441201722
1.96 St Dev	0.527128507	0.543581313	0.556450345	0.666735212	0.728111172	0.864755375

	RE P-1 : P-2	RE P-2 : P-3	RE P-1 : P-3	LE P-1 : P-2	LE P-2 : P-3	LE P-1 : P-3
Mean	0.013809229	-0.003280064	0.010529166	-0.022516799	0.011416734	-0.011100065
St Dev	0.153529883	0.221433106	0.214333508	0.190495559	0.200546072	0.217590792
1.96 St Dev	0.300918571	0.434008888	0.420093675	0.373371296	0.393070301	0.426477952