

RESULTS

Quantitative Findings

To answer the first research question: What are the differences in higher education female students' academic achievement between students in the flipped classrooms and students in traditional classrooms in the Computer in Education course, the results from the statistical analyses, more specifically from the descriptive statistics, repeated-measures ANOVA and

Analysis of Covariance are presented in this section. A repeated-measures ANOVA for assessing the mean growth in scores for the two groups is presented as well as an analysis for covariance ANCOVA for assessing whether there were mean differences in the posttest scores, controlling for the pretest scores.

Descriptive Statistics

The study sample included 28 students. Each group included 14 students. Table 2 shows the mean pre-test score is $M = 14.32$ ($SD = 4.13$) and the mean post-test score is $M = 18.36$ ($SD = 3.33$). Table A.a also provides the descriptive statistics by group. For the traditional group, the mean pre-test score is $M = 14.71$ ($SD = 4.12$), and the mean post-test score is $M = 17.14$ ($SD = 2.25$), while for the flipped group, the mean pre-test score is $M = 13.93$ ($SD = 4.25$) and the mean post-test score is $M = 19.57$ ($SD = 3.84$). Appendices A.b and A.c provides a more complete set of descriptive statistics for the sample.

Table 2

Descriptive Statistics (n , M , SD) for Both Groups, and the Total Sample

	<i>Tradition</i>			<i>Flipped</i>			<i>Total</i>		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Pretest	14	14.71	4.122	14	13.93	4.251	28	14.32	4.128
Posttest	14	17.14	2.248	14	19.57	3.837	28	18.36	3.325

The box plots in Figures 2 and 3 correspond to pre- posttest scores for the entire sample and shows some outliers for the pre-test scores. Figures 4 and 5 provide the boxplots by group.

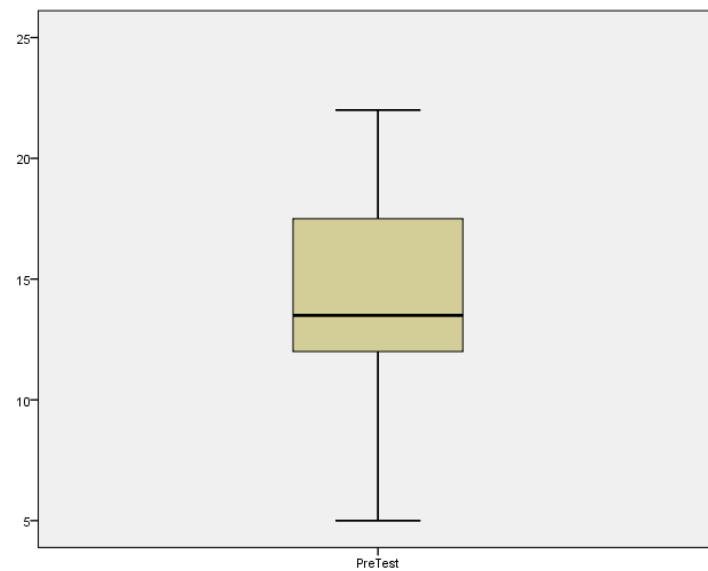


Figure 2: Box plot of pre-test scores.

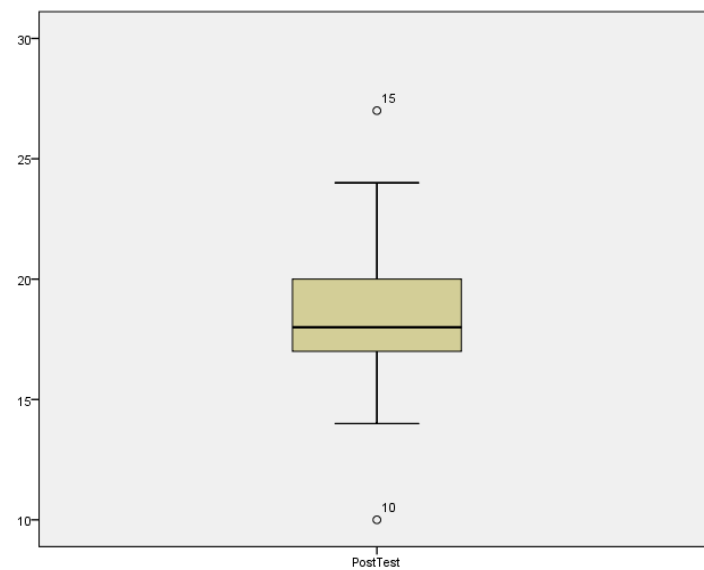


Figure 3: Box plot of post-test scores.

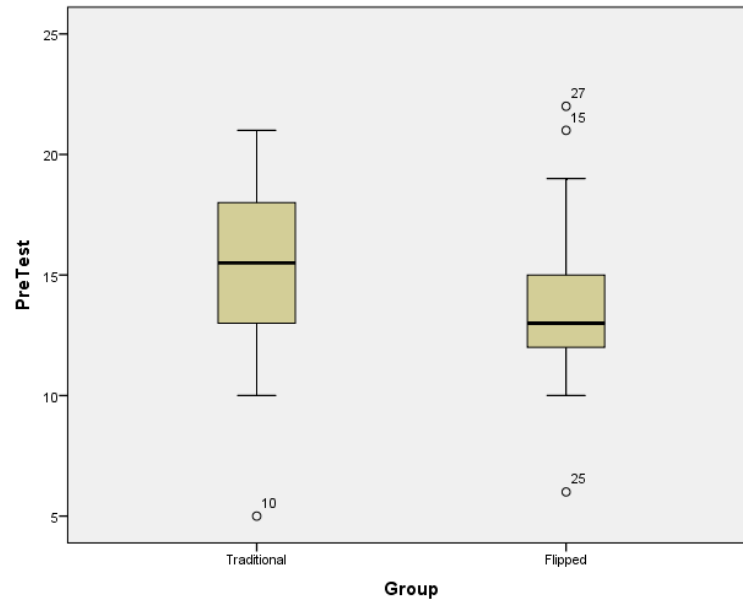


Figure 4: Box plot of pre-test scores by group.

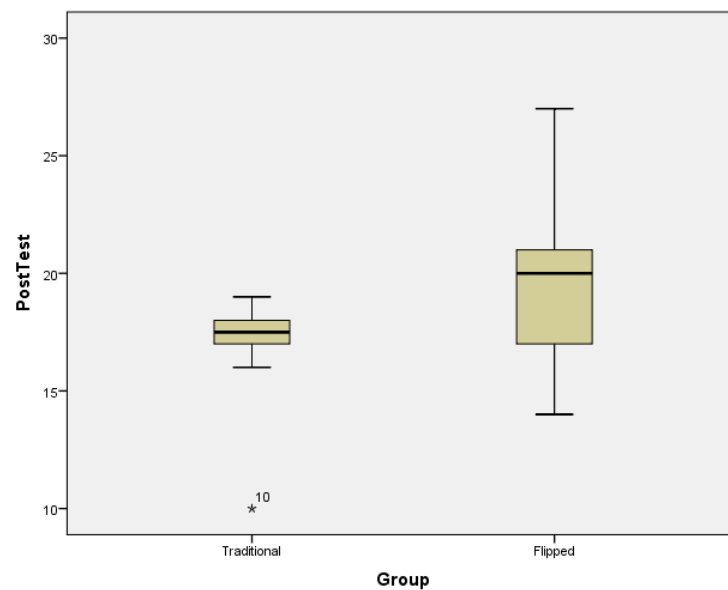


Figure 5: Box plot of post-test scores by group.

It is clear that the variability of the posttest is much larger for the flipped group than for the traditional group.

Repeated Measures ANOVA

The plot in Figure 6 shows the marginal means by time for both the traditional and flipped groups.

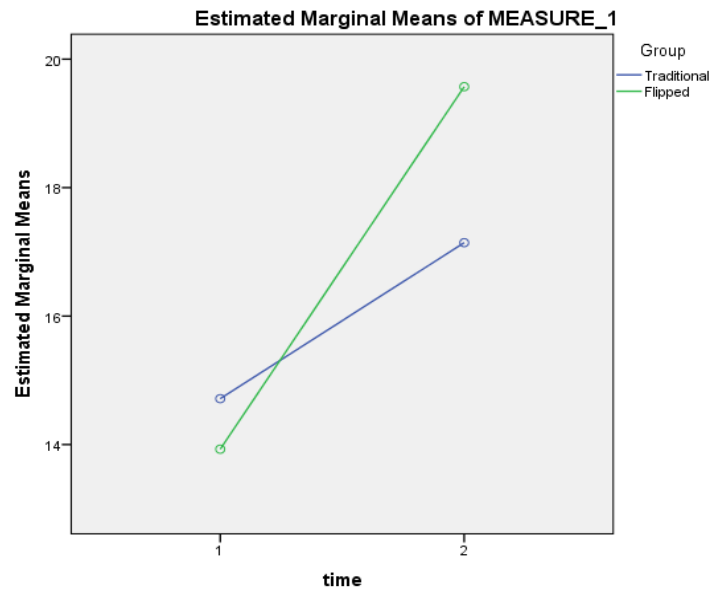


Figure 6: Marginal means by group.

At the sample level, the marginal mean plots intersect, which suggests the mean growth in scores may differ for the two groups. Appendix A.d shows the assumption of homogeneity of covariance is met, $F(3, 121680) = 2.540, p = .055$.

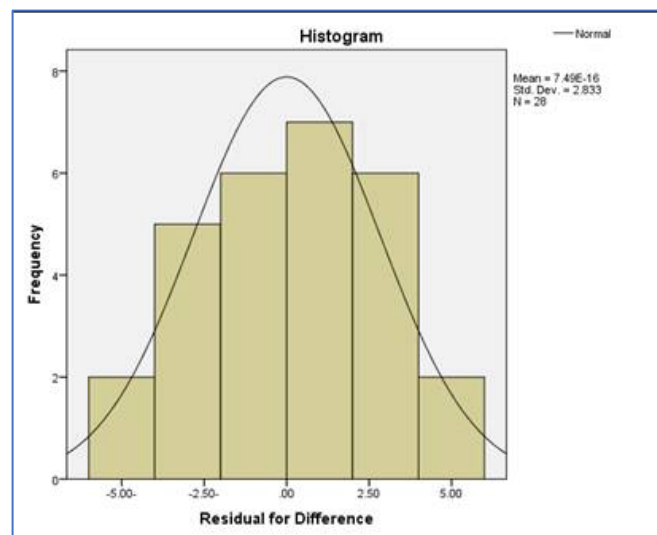


Figure 7: Residuals of difference scores.

Also, the histogram of the residuals for the difference scores (see Figure 7) do not seem to depart significantly from normality, which is confirmed by the results found in normality test results shown in Appendix A.e, with $KW = .09$, $p = .200$. Appendix A.f also shows the results of Levene's test for the score differences, with $F(1, 26) = 0.50$, $p = .488$, and together with the fact that the ratio of variances, at 1.54, is less than 2.0, and therefore it is concluded the homogeneity of variances is met.

The results of the repeated-measures ANOVA are shown in Appendix A.g. The effect of time is statistically significant, $F(1, 26) = 54.73$, $p < .001$, and there is an overall increase from pre to post scores. Additionally, the change in scores depended significantly on the group, as the interaction Time \times Group is significant, $F(1, 26) = 8.68$, $p = .007$, with an effect size of $\eta^2 = 0.06$, which indicates a moderate effect size (Cohen, 1988). Appendix A.h. The p-value of $p = .007$ indicates a probability of 0.007 of seeing a sample difference in growth as extreme or more extreme the one observed, under the assumption that the difference in growth between groups is not present in the population.

Analysis of Covariance

The assumptions for ANCOVA are normality, homogeneity of variances, and homogeneity of regression slopes.

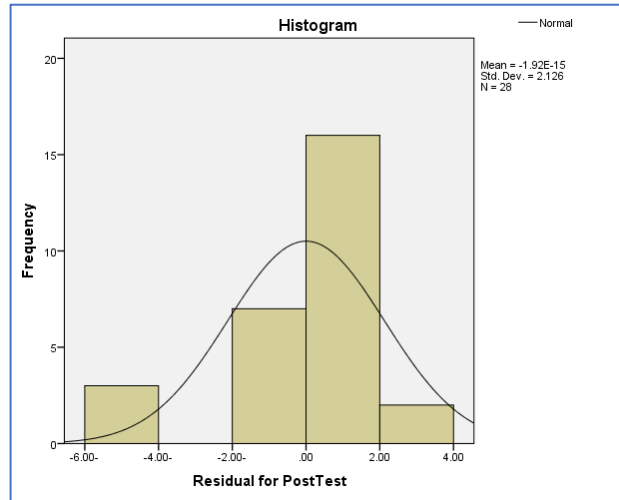


Figure 8: Residuals for the ANCOVA analysis.

Figure 8 shows that the distribution of residuals is left-skewed, which suggests that normality is violated, but the results of the Kolmogorov-Smirnov test shown in Appendix A.i indicates that this departure from normality was not statistically significant, $KW = 0.15$, $p = .112$, although the results of the Shapiro-Wilk test suggest that significant non-normality is indeed evident ($SW = 0.92$, $p = .036$).

Appendix A.j shows that the homogeneity of regression slopes assumption is met, $t(24) = .83$, $p = .415$, but Appendix A.k shows that the homogeneity of variances assumption is not met, $F(1, 26) = 6.14$, $p = .020$. Additionally, the pretest was not a significant predictor of posttest, $t(24) = .85$, $p = .404$.

The results from the ANCOVA in Appendix A.k show a statistically significant group effect on the posttest scores after controlling for the pretest scores, $F(1, 25) = 11.59$, $p = .002$, with an effect size of $\eta^2 = 0.19$, which indicates a large effect size (Cohen, 1988). The adjusted marginal mean post-test score for the traditional group is 16.93, and the adjusted marginal mean posttest score for the flipped group is 19.79 (see Figure 9).

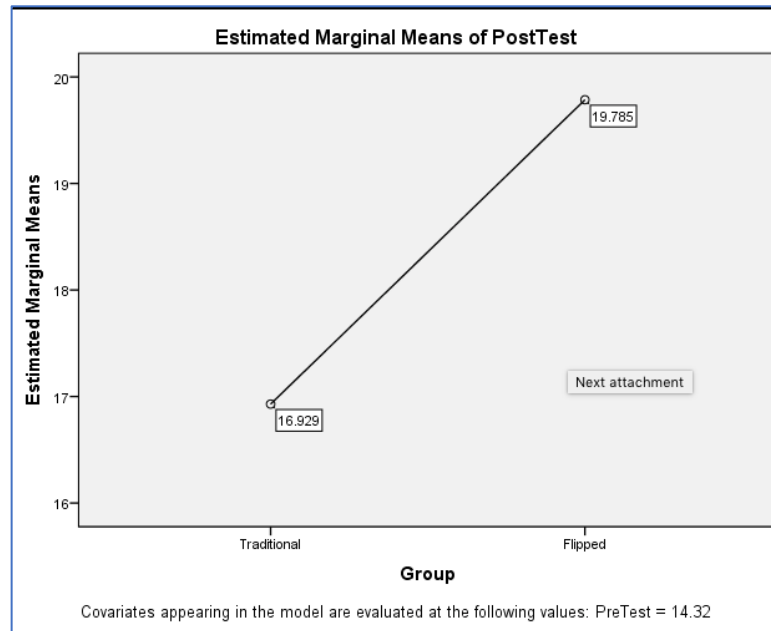


Figure 9: Adjusted marginal means.

Summary of Quantitative Findings

From the descriptive analysis, the variability of the posttest was much larger for the flipped group than for the traditional group. Moreover, from the repeated measures analysis, it can be concluded the interaction between time and group was significant, which was reflected in the change in mean from time 1 to time 2 was significantly larger for the flipped group than for the traditional group. On the other hand, the ANCOVA found the effect on the group was significant on the posttest scores, even after controlling for the pretest scores. This leads to the conclusion that there were statistically significant differences in higher education female students' academic achievement between students in the flipped classrooms and students in

traditional classrooms in the Computer in Education course that favored students in the flipped classrooms.

APPENDIX A

QUANTITATIVE RESULTS

A.a Descriptive Statistics				
			Statistic	Std. Error
PreTest	Mean		14.32	.780
	95% Confidence Interval for	Lower Bound	12.72	
	Mean	Upper Bound	15.92	
	5% Trimmed Mean		14.41	
	Median		13.50	
	Variance		17.041	
	Std. Deviation		4.128	
	Minimum		5	
	Maximum		22	
	Range		17	
	Interquartile Range		6	
	Skewness		-.165	.441
	Kurtosis		.119	.858
	Mean		18.36	.628
	95% Confidence Interval for	Lower Bound	17.07	
PostTest	Mean	Upper Bound	19.65	
	5% Trimmed Mean		18.33	
	Median		18.00	
	Variance		11.053	
	Std. Deviation		3.325	
	Minimum		10	
	Maximum		27	
	Range		17	
	Interquartile Range		3	
	Skewness		.262	.441
	Kurtosis		1.750	.858

A.b Descriptive Statistics of Pre-Test score by Group				
Group		Statistic		Std. Error
PreTest	Traditional	Mean	14.71	1.102
		95% Confidence Interval for Lower Bound	12.33	
		Mean Upper Bound	17.09	
		5% Trimmed Mean	14.90	
		Median	15.50	
		Variance	16.989	
		Std. Deviation	4.122	
		Minimum	5	
		Maximum	21	
		Range	16	
		Interquartile Range	6	
		Skewness	-.900	.597
		Kurtosis	1.067	1.154
	Flipped	Mean	13.93	1.136
		95% Confidence Interval for Lower Bound	11.47	
		Mean Upper Bound	16.38	
		5% Trimmed Mean	13.92	
		Median	13.00	
		Variance	18.071	
		Std. Deviation	4.251	
		Minimum	6	
		Maximum	22	
		Range	16	
		Interquartile Range	4	
		Skewness	.499	.597
		Kurtosis	.522	1.154

A.c Descriptive Statistics Post-Test score by Group				
Group			Statistic	Std. Error
PostTest	Traditional	Mean	17.14	.601
		95% Confidence Interval for Mean		
		Lower Bound	15.84	
		Upper Bound	18.44	
		5% Trimmed Mean	17.44	
		Median	17.50	
		Variance	5.055	
		Std. Deviation	2.248	
		Minimum	10	
		Maximum	19	
	Flipped	Range	9	
		Interquartile Range	1	
		Skewness	-2.718	.597
		Kurtosis	8.815	1.154
		Mean	19.57	1.026
		95% Confidence Interval for Mean		
		Lower Bound	17.36	
		Upper Bound	21.79	
		5% Trimmed Mean	19.47	
		Median	20.00	
		Variance	14.725	
		Std. Deviation	3.837	
		Minimum	14	
		Maximum	27	
		Range	13	
		Interquartile Range	5	
		Skewness	.202	.597
		Kurtosis	-.321	1.154

A.d Box's Test of Equality of Covariance Matrices^a

Box's M	8.313
F	2.540
df1	3
df2	121680.000
Sig.	.055

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

a. Design: Intercept + Group

Within Subjects Design: time

A.e Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual for Difference	.086	28	.200*	.980	28	.858

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

A.f Levene's Test of Equality of Error Variances^a

Dependent Variable: Difference

F	df1	df2	Sig.
.495	1	26	.488

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

a. Design: Intercept + Group

A.g Tests of Within-Subjects Effects							
Measure: MEASURE_1							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
time	Sphericity Assumed	228.018	1	228.018	54.730	.000	.678
	Greenhouse-Geisser	228.018	1.000	228.018	54.730	.000	.678
	Huynh-Feldt	228.018	1.000	228.018	54.730	.000	.678
	Lower-bound	228.018	1.000	228.018	54.730	.000	.678
time *	Sphericity Assumed	36.161	1	36.161	8.680	.007	.250
	Greenhouse-Geisser	36.161	1.000	36.161	8.680	.007	.250
	Huynh-Feldt	36.161	1.000	36.161	8.680	.007	.250
	Lower-bound	36.161	1.000	36.161	8.680	.007	.250
Group	Sphericity Assumed	108.321	26	4.166			
	Greenhouse-Geisser	108.321	26.000	4.166			
	Huynh-Feldt	108.321	26.000	4.166			
	Lower-bound	108.321	26.000	4.166			
Tests of Within-Subjects Contrasts							
Measure: MEASURE_1							
Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	
time	Linear	228.018	1	228.018	54.730	.000	
Time* Group	Linear	36.161	1.000	36.161	8.680	.007	
Error (time)	Linear	108.321	26.000	4.166			
Tests of Between-Subjects Effects							
Measure: MEASURE_1							
Transformed Variable: Average							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	
Intercept		14950.4	1	14950.4	642.916	.000	
Group		9.446	1	9.446	.406	.529	
Error		604.607	26	23.254			

A.h Tests of Normality for ANCOVA

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual for PostTest	.149	28	.112	.921	28	.036

a. Lilliefors Significance Correction

A.i Regression Coefficients

Model	Unstandardized		Standardized	t	Sig.
	Coefficients		Coefficients		
	B	Std. Error	Beta		
1 (Constant)	10.063	5.023		2.003	.057
Group	.380	3.104	.058	.122	.904
PreTest	.283	.333	.351	.850	.404
PreTest_x_Group	.173	.208	.487	.829	.415

a. Dependent Variable: PostTest

A.j Levene's Test of Equality of Error Variances^a

Dependent Variable: PostTest			
F	df1	df2	Sig.
6.143	1	26	.020

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

a. Design: Intercept + PreTest + Group

A.k. Tests of Between-Subjects Effects

Dependent Variable: PostTest						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	176.382 ^a	2	88.191	18.065	.000	.591
Intercept	229.593	1	229.593	47.030	.000	.653
PreTest	135.096	1	135.096	27.673	.000	.525
Group	56.574	1	56.574	11.589	.002	.317
Error	122.047	25	4.882			
Total	9734.000	28				
Corrected Total	298.429	27				

a. R Squared = .591 (Adjusted R Squared = .558)

