Honors Theses

The Division of Undergraduate Studies

2012

AP Student Visual Preferences for Problem Solving

Liesl Swoyer

Abstract

The purpose of this study is to explore the mathematical preference of high school AP Calculus students by examining their tendencies for using differing methods of thought. A student's preferred mode of thinking was measured on a scale ranging from a preference for analytical thought to a preference for visual thought as they completed derivative and antiderivative tasks presented both algebraically and graphically. This relates to previous studies by continuing to analyze the factors that have been found to mediate the students' performance and preference in regards to a variety of calculus tasks. Data was collected by Dr. Erhan Haciomeroglu at the University of Central Florida. Students' preferences were not affected by gender. Students were found to approach graphical and algebraic tasks similarly, without any significant change with regards to derivative or antiderivative nature of the tasks. Highly analytic and highly visual students revealed the same proportion of change in visuality as harmonic students when more difficult calculus tasks were encountered. Thus, a strong preference for visual thinking when completing algebraic tasks was not the determining factor of their preferred method of thinking when approaching graphical tasks.

Keywords: Calculus, Derivative, Antiderivative, Gender Difference, Visuality, Visual Thinking, Analytic Thinking, Harmonic Thinking, Preference, High School.

THE FLORIDA STATE UNIVERSITY

DEPARTMENT OF STATISTICS

AP STUDENT VISUAL PREFERENCES FOR PROBLEM SOLVING

By

LIESL A. SWOYER

A Thesis submitted to the

Department of Statistics

in partial fulfillment of the requirements for graduation

with Honors in the Major

Degree Awarded:

Spring Semester, 2012

The members of the Defense Committee approve the thesis of Liesl A. Swoyer defended on March 21, 2012.

Dr. Eric Chicken
Thesis Director

Dr. Daniel McGee
Committee Member

Dr. Steve Paris
Outside Committee Member

Introduction

Previous studies that examine the relationship of preferred method of thinking with mathematical performance have produced mixed results. The inventors of the Mathematical Processing Instrument for Calculus, or the MPIC, [1] hypothesized that the differences in mathematical performance might be caused by students' underlying preference for specific methods of thought, and thus designed the MPIC to reveal what each student's preferred method was. They believe that a student's ability to master calculus is directly related to their ability to harmonize analytical and visual methods of thought in their approach to a graphically presented antiderivative or derivative task. They believe that by clearly defining preference as a separate entity from ability, they may better explore the mediating factors affecting mathematical performance in calculus based courses.

The approach taken by this current study has built upon the findings of many other studies as interest in the relationships between preference and performance has been well established. The study performed by Carr, Steiner, Kyser, and Biddlecomb [2] with second grade students compared to the study by Orhun [3] with college students revealed that while mathematical competency and the use of cognitive strategy was more prevalent with second grade boys over girls, there existed no significant difference in mathematical achievement between men and women at the college level. The only difference that Orhun discovered was that there existed a significant difference in the learning methods employed by the male and female college students that he studied.

Fennema, Carpenter, Jacobs, Franke, and Levi [4], as a result of their study of elementary students, found that female students very often used concrete strategies when solving mathematical problems while boys tended to employ abstract strategies to solve problems. The findings of that study are a contradiction to the findings of Lowrie and Kay [5] in their study of middle school students in the sixth grade, because these researchers saw no significant difference in visual preference, whether it be visual or analytical methods of thought, across gender. Again, the findings are in contradiction of each other.

Krutetskii [6] hypothesized that mathematical performance was a result of a student's capacity to apply analytical methods of thought, while the strength of their aptitude to implement visual methods of thought determined the kind of mathematical ability the student possessed. Moses [7] and Suwarsono [8] researched the mathematical abilities of elementary students and middle school students. Their studies did not reveal a correlation between visual aptitude and mathematical performance, either. Contradictory to the findings of these mathematicians, Ferrini-Mundy [9] and Ubuz [10] in fact found that there was a very distinctive relationship between visual proficiency and mathematical performance.

Again, it is the belief of the inventors of the Mathematical Processing Instrument for Calculus (MPIC) that some preference for a method of thought does not directly relate to a student's ability to use other methods of thought. This distinction between preference and ability has been ignored in the reviewed aforementioned research studies. Preference is not necessarily the same as ability, although these visual preferences may be affecting their mathematical ability. This continuation of the study conducted by the inventors of the MPIC sought to further explore visual preference using the MPIC and asked the following questions:

1. Do students change their preference when presented with graphic, algebraic, derivative, or antiderivative tasks?

2. Do students differ in visuality on graphic derivative, visuality on algebraic derivative, visuality on graphic antiderivative, and visuality on algebraic derivative tasks?

3. Do high-visual and high-analytic students change their preference when presented with more difficult derivative-antiderivative tasks?

Methodology

Participants

All of the data used was provided by Dr. Erhan Haciomeroglu of the School of Teaching, Learning, and Leadership at The University of Central Florida. Any identifying information had already been removed by Dr. Haciomeroglu. I only had access to student genders, scores on mathematical performance exams and scores from visuality assessment tools. Data was collected from AP Calculus students attending high school in various school districts in Northern and Central Florida.

Procedure

Students were given calculus derivative and antiderivative tasks presented graphically. The Mathematical Processing Instrument for Calculus (MPIC) determined the student's preference for visual or analytical thinking. A questionnaire was given to each student after completing all the tasks provided by the MPIC. The questionnaire asked the students to select a method of solution, either visual or analytical, that most closely resembled their own. Select students were interviewed by the data collectors to give additional insight into their visual preferences.

Analysis

With the intent of exploring the data associated with the visualities of AP Calculus students currently in high school in North Florida, I ran several analyses. In the study's first question, the area of interest was discerning whether or not students have different preferences when presented with graphic, algebraic, derivative, or antiderivative tasks. This analysis was performed by examining students' visualities measured in graphic (calculus) derivative (DGV), algebraic derivative (DAV), graphic antiderivative (AGV), and algebraic antiderivative (AAV) test questions. It is assumed that there are a percentage of highly visual and highly analytic students within the data set. After analyzing multiple comparisons of student visualities, separate classifications were made categorizing students as 'Highly Visual' and 'Highly Analytic'. Creating these two subsets addressed question two of the study, which was to determine whether a highly analytic or visual student will change their preferred method of solving calculus problems when faced with more difficult derivative and antiderivative tasks.

For each student, the data consists of one or two AP scores, twenty visual preference scores from the MPIC test questions. These MPIC visual test questions were each divided into seven graphical derivative, seven algebraic derivative, three algebraic antiderivative, and three graphical antiderivative questions. The visuality data also included interview results for select students which were used in previous papers to validate the MPIC questions. Additionally, I know the gender of each student. This data has already been collected.

The analysis began by first reducing the complexity of the data through the creation of statistics for each student. I used mean scores for MPIC preference on the seven graphical derivative, seven algebraic derivative, three algebraic antiderivative, and three graphical antiderivative questions, and the quartiles for MPIC visuality scores. Average preference scores were further assembled into scores that involved derivatives or antiderivatives. The quartiles were used to designate students as visual (top quartile of the students with respect to MPIC visuality), analytic (bottom quartile) or harmonic (any remaining students). Following this, exploratory data analysis was used to get a feel for the properties of the data.

Initially, I used an analysis of variance (ANOVA) [11] to see if performance scores were affected by gender and visuality. I used both parametric and nonparametric methods since prior research shows that this type of data does not necessarily meet the strict assumptions required for parametric tests. Parametric ANOVA is based on normality assumptions for the errors within the data, while nonparametric ANOVA will not rely on this assumption. Both types of methods assume constant variance. I use Kruskal-Wallis analysis and permutation methods for the nonparametric ANOVA [12, 13, 14]. The permutation method is computationally intensive, but is valid over a wider range of assumptions than ANOVA.

Variables

To analyze changes of preference among students in graphic, algebraic, derivative, and antiderivative tasks, I created new variables. A matrix of the student responses that determine their visuality in the four task groups of interest was constructed. It combined graphical antiderivative (AGV), graphical derivative (DGV), algebraic antiderivative (AAV), and algebraic derivative (DAV) visuality scores from all of the students in addition to the unweighted mean (CV_U) and weighted mean (CV_W) of the students' visualities. The weighted mean was calculated by taking into consideration the number of questions in each category and weighting their averages accordingly. All tests conducted use a significance level of 0.05.

Results

An ANOVA was conducted to explore the students' changes in preference when presented with graphic, algebraic, derivative, and antiderivative tasks. The means for all 150 students in the study for the calculus visualities are referenced in Figure 1 and Table 1.

Variable	*Mean*
CV_U	0.95
DGV	1.15
DAV	0.61
AGV	1.04
AAV	0.59
CV_W	0.85

Table 1. Means of Calculus Visualities

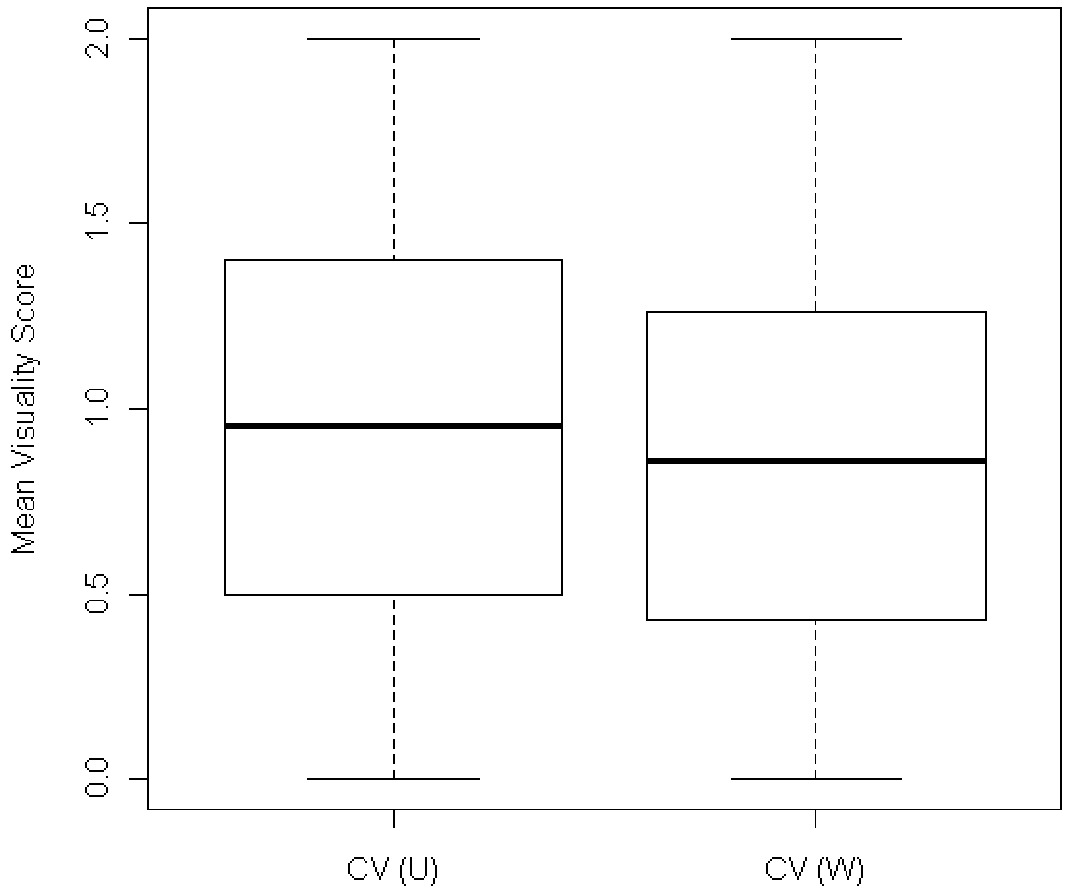

Figure 1. Box-and-Whisker Plot of the Weighted (CV_W) and Unweighted (CV_U) Means

I created an ANOVA containing the variables gender and treatment. I ran an analysis using treatment, gender, and their interaction as the factor levels. Treatment included the graphical antiderivative, graphical derivative, algebraic antiderivative, and algebraic derivative questions from the MPIC. According to the output below, it did not appear that gender or the interaction of gender and treatment significantly affected the data; however, treatment was significant.

Response: Visuality

	Sum Sq	Df	F value	p-value
Treatment	36.844	3	23.6522	0.0000
Gender	1.268	1	2.4415	0.1187
Treatment:Gender	3.078	3	1.9759	0.1164
Residuals	307.392	592		

Table 2. ANOVA using Treatment, Gender, and their interactions a factor

I used the Shapiro-Wilks test to check for normality [15]. It appeared that the data was not normal; see the Q-Q plot in Figure 2. The test statistic was 0.9537 and the p-value was small (p=0.0000); however, the sample size was large so I felt comfortable proceeding with the ANOVA.

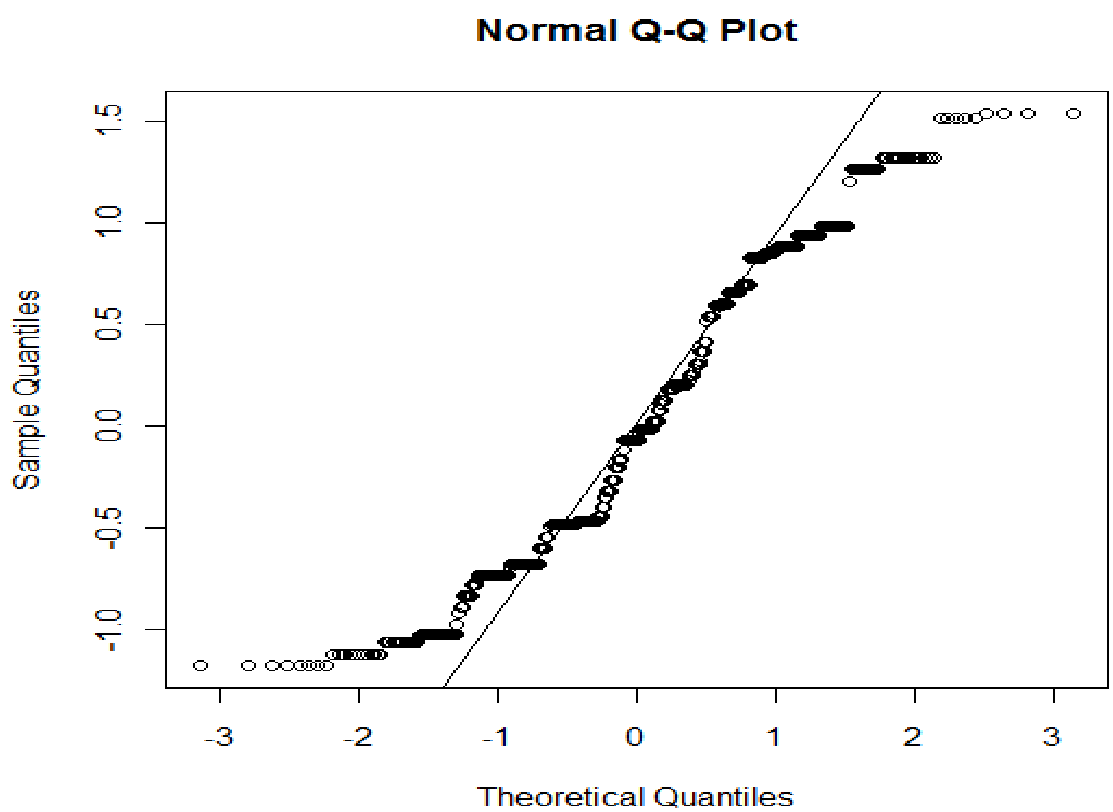

Figure 2. Q-Q Plot for Normality

Next, I conducted Bartlett's Test for the homogeneity of variances [16]. Bartlett's test determined that the data did not have significantly different variances since the Bartlett K-square was 3.6841 and the p-value was 0.2977. I did not reject the null hypothesis that the variances were equal. However, Bartlett's test is sensitive to non-normal data, so I ran Levene's Test, which has no assumption of normality, to check for homogeneity of variances [17]. The test statistic was 2.4893 on 7 degrees of freedom, with a p-value of 0.0159. This implies that the variances are not all the same, not the same conclusion as Bartlett's test. Since the data is not normal, Bartlett's test may be failing. I will go with the results of Levene's test.

The data was shown to have unequal variances. Since the assumption of equal variances was not met, I decided to conduct a weighted ANOVA on the data (Table 4). Again, I believe large sample size will make non-normality a non-issue.

Response: Visuality

	Sum Sq	Df	F value	p-value
Treatment	52.92	3	24.5392	0.0000
Gender	2.26	1	3.1480	0.0765
Treatment:Gender	3.28	3	1.5195	0.2083
Residuals	425.52	592		

Table 4. Weighted ANOVA on Visuality

The weighted ANOVA agreed with the unweighted ANOVA in that treatment was a significant factor while gender and the interaction factor were not (Table 4). For confirmation, I decided to conduct a nonparametric permutation version of ANOVA on the three factors (Table 5).

	Df	SumsOfSqs	MeanSqs	F.Model	p-value
Treatment	3	36.84	12.2812	23.6008	0.0010
Gender	1	1.39	1.3861	2.6637	0.1030
Treatment:Gender	3	2.29	0.7634	1.4670	0.2220
Residuals	592	308.06	0.5204		
Total	599	348.58			

Table 5. Permutations Test on Visuality

Again, the permutations test agreed with the ANOVA, even though this test did not take into account differing variances and hence the p-values may not be strictly accurate. The students seemed to change their preference for analytical or visual methods when presented with different treatments.

Because of non-normality, I also ran a Kruskal-Wallis rank sum test on the visuality scores for the eight treatments formed by the combination of treatment on both genders [18]. The treatments, in order, are graphical derivative (DG) for males and females, algebraic derivative (DA) for males and females, graphical antiderivative (AG) for males and females, and algebraic antiderivative (AA) for males and females. Due to the large size of the output from this test, Table 23a and Table 23b can be seen in the Appendix.

The output from the test suggests that a significant difference in visuality occurs when females encounter graphical derivative and algebraic derivative questions. The results also show that males approach graphical derivative and algebraic derivative questions differently. Female visuality scores for graphical antiderivative and algebraic antiderivative questions differ significantly. For females, it is apparent that the nature of the problem, whether graphical or algebraic, significantly affects their visuality score.

The visuality score of females for graphical derivative questions is also different from the visuality score of males for algebraic derivative questions. It remains unknown whether the change is a result of gender or question type. A similar problem is found with the comparison of female visuality scores for algebraic antiderivative and graphical derivative questions. It is apparent that females on average approach these two questions differently; however, it is unapparent as so whether that change is a result of the transition from graphical to algebraic or from derivative to antiderivative. Males approach algebraic antiderivative questions differently than females approach graphical derivative questions. The reason for this change could be because of gender, graphical/algebraic, or derivative/antiderivative transitions. Male visuality scores for graphical derivative questions are significantly different from female visuality scores for algebraic derivative questions. This difference could be the result of gender or subject change from graphical to algebraic.

As there are 28 total comparisons, the output from the Kruskal-Wallis rank sum test became difficult to interpret. Many of the comparisons with significant differences had no conclusive findings as to the cause of their difference. This is why I prefer to use the weighted ANOVA results.

Although the assumption of normality was not met, the sample size was large; hence, multiple comparisons for the pairwise differences were conducted on the combination of treatment and gender so as to compare the ANOVA output with Kruskal-Wallis. Table 7 for the ANOVA can also be referenced in the Appendix. The method used is Tukey's HSD [19].

The significant differences correspond with the output from the Kruskal-Wallis rank sum test with the exception of one. I am satisfied that the Kruskal-Wallis rank sum test was accurate.

The interaction between gender and treatment is not significant, so a reduced model for both the Kruskal-Wallis and the ANOVA was conducted to analyze the factors treatment and gender without their interaction term.

I also ran an ANOVA with gender as the only factor to determine whether or not one gender of students changed their preference significantly more or less than the other gender. Although the initial model showed gender was not a significant factor, I ran this to confirm this conclusion without any possible effect due to including the treatment factor. These tests can be referenced in the Appendix (Table 24a and Table 24b). No new conclusions were drawn.

I then conducted a one-way ANOVA on the reduced model, which accounted for the four treatment factors (Table 6). The response continues to be visuality scores measured through CV_W.

	Sum Sq	Df	F value	p-value
Treatment	36.844	3	23.48	0.0000
Residuals	311.738	596		

Table 6. ANOVA table for visuality on treatment (DGV, AGV, DAV, AAV)

I followed-up with a Tukey HSD test to determine which variables have significantly different means (Table 7).

	diff	lwr	upr	p adj
2-1	**-0.5333333**	**-0.7484808**	**-0.3181859**	**0.0000000**
3-1	-0.1076190	-0.3227665	0.1075284	0.5704962
4-1	**-0.5533333**	**-0.7684808**	**-0.3381859**	**0.0000000**
3-2	**0.4257143**	**0.2105668**	**0.6408617**	**0.0000028**
4-2	-0.0200000	-0.2351474	0.1951474	0.9951655
4-3	**-0.4457143**	**-0.6608617**	**-0.2305668**	**0.0000008**

Table 7. Tukey HSD Test for Equality of Means. Bold are significant differences.

The data shows there exists a significant difference in the means of graphical and algebraic derivative visuality, graphical derivative and algebraic antiderivative visuality, algebraic derivative and graphical antiderivative visuality, and graphical antiderivative and algebraic antiderivative visuality. The data did not suggest that there was any significant difference between graphical derivative and graphical antiderivative visuality, nor between algebraic derivative and algebraic antiderivative visuality.

In short, the differences found by the Tukey HSD test were a result of the transition from graphical to algebraic questions or vice versa. No significant change of preference was found between antiderivative and derivative questions.

I was not convinced that the data satisfied the assumptions of normality and equality of variances for the ANOVA, so I first conducted the Shapiro-Wilks test for normality. With a test statistic of 0.9372 and an extremely small p-value (p=0.0000) the data was not normal (Figure 3).

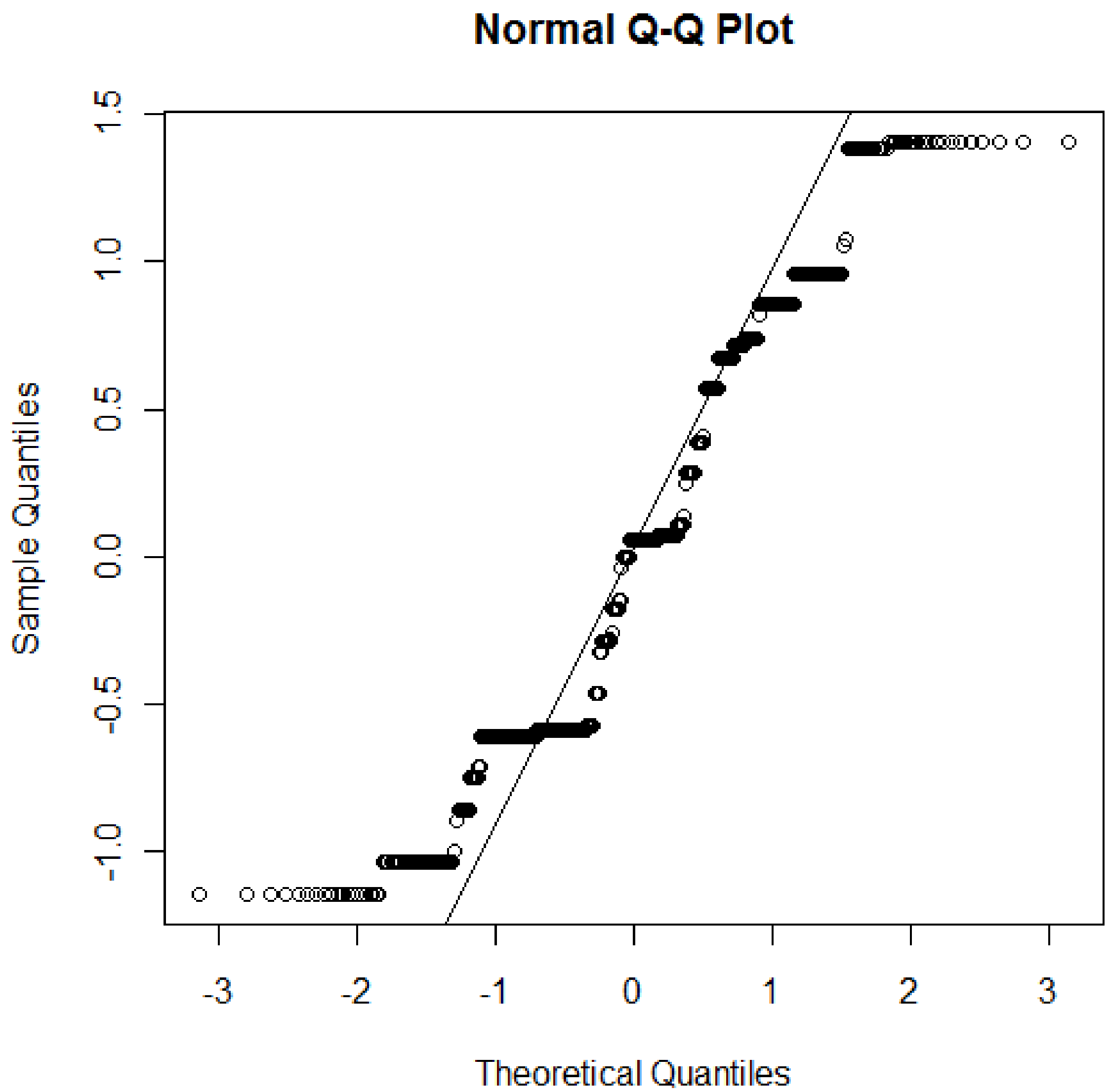

Figure 3. QQ Plot for Normality

I next conducted the Bartlett Test of homogeneity of variances. The Bartlett K-squared statistic was 3.6841 and the p-value was 0.2977. I did not reject the null hypothesis that there was no difference in the variances. The Bartlett Test, however, is susceptible to non-normal data; thus, I conducted Levene's Test, which is not susceptible to non-normal data. Levene's Test yielded a p-value of 0.0043 and a test statistic of 4.4347, which did not agree with the Bartlett Test. The variances are not equal, see Figure 4. It can be seen that the Bartlett Test failed as a result of non-normal data.

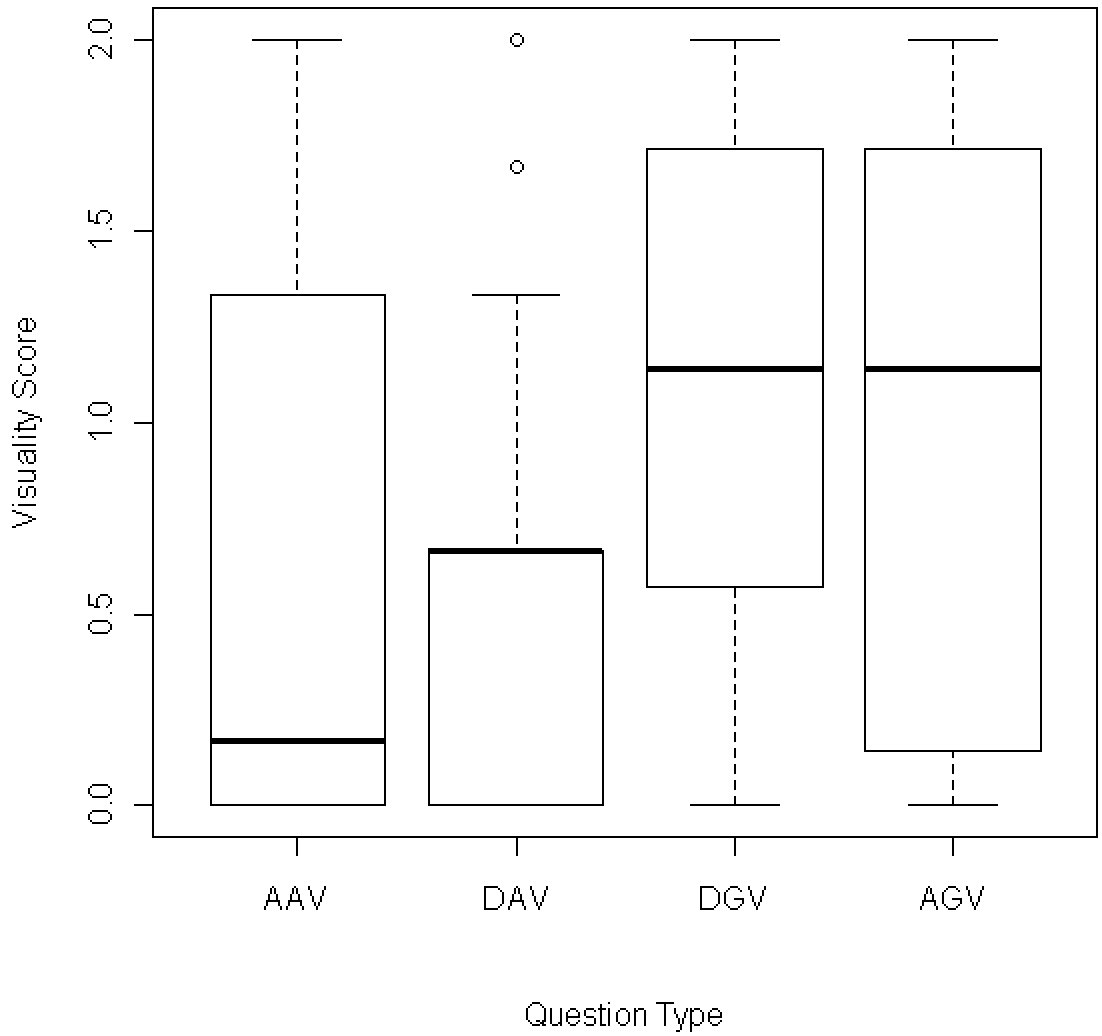

Figure 4. Box-and-Whisker Plot of Visuality scores for all 150 students in the study.

The sample size was large, so I proceeded with the parametric ANOVA. Since the data had unequal variances, I used a weighted ANOVA on the reduced model. Treatment was found to be significant with a p-value of 0.0000.

A nonparametric permutations test was then run on the reduced model. The permutations test had a p-value of 0.0010 for treatment, which was in agreement with both the unweighted and weighted ANOVA for the reduced model (Table 8). Note that the permutation test did not account for non-constant variance.

	Df	SumsOfSqs	MeanSqs	F.Model	p-value
Treatment	3	36.84	12.281	23.48	0.0010
Residuals	596	311.74	0.523		
Total	599	348.58			

Table 8. Nonparametric Permutations Test

I decided to conduct the Multiple Comparison Test after Kruskal-Wallis and the Holm-Bonferroni method [20]. The p-value (p=0.0000) for the Kruskal-Wallis Rank Sum test was consistent with the p-value for the ANOVA. The output of the Multiple Comparison Test after Kruskal-Wallis (Table 9) was also consistent with the Holm-Bonferroni adjustment tests (Table 10). I was satisfied that the results were accurate.

	obs.dif	critical.dif	difference
1-2	**117.910000**	**52.8091**	**TRUE**
1-3	26.896667	52.8091	FALSE
1-4	**126.446667**	**52.8091**	**TRUE**
2-3	**91.013333**	**52.8091**	**TRUE**
2-4	8.536667	52.8091	FALSE
3-4	**99.550000**	**52.8091**	**TRUE**

Table 9. Multiple Comparison Test After Kruskal-Wallis

trt 1	trt 2	p	Bon	rej=1	Holm	rej=1
1	2	0.00000	0.00833	1	0.008333333	1
1	4	0.00000	0.00833	1	0.010000000	1
2	3	0.00000	0.00833	1	0.012500000	1
3	4	0.00000	0.00833	1	0.016666667	1
1	3	0.21514	0.00833	0	0.025000000	0
2	4	0.80340	0.00833	0	0.050000000	0

Table 10. Holm and Bonferroni adjustments for unequal varaince t-tests (trt=treatment; rej=rejection).

So, the reduced model using visuality as response and treatment as factor completely agrees with the full model.

Since the previous analyses had shown that students used similar methods when presented with graphical antiderivative and graphical derivative questions, I pooled those treatments into one group so as to compare them with the algebraic derivative and algebraic antiderivative questions, which were also combined into a single treatment group. I conducted an ANOVA on the range of visuality scores given on graphical and algebraic questions while again taking into account gender as a factor. I created two new variables to combine responses in treatment and simplify the data into responses of one or two. The ANOVA output revealed that treatment was significant and that, again, gender was an insignificant factor (Table 11). There were no new conclusions. (The interaction term was also insignificant; the results shown are a reduced model.)

Response: Visuality

	Sum Sq	Df	F value	p-value
Treatment	35.945	1	68.9189	0.0000
Gender	1.268	1	2.4306	0.1195
Residuals	311.369	597		

Table 11. ANOVA using treatment and gender as factors

The difference in mean visuality for algebraic questions is 0.48 below the mean visuality for graphical (calculus) questions. The Shapiro-Wilks test reconfirmed that the data was not normal with a test statistic of 0.9439 and a small p-value (p=0.0000). I then conducted the Bartlett test for homogeneity of variances. The Bartlett test statistic was 6.6293 and the p-value was 0.0847. Knowing that the Bartlett Test was sensitive, I ran Levene's Test as well. Levene's Test resulted in a test statistic of 4.9358 and a p-value of 0.0022, and thus did not agree with Bartlett's Test. I use Levene's Test as it is not sensitive to non-normality. The data does not have constant variance. This test implies non-constant variance and so weighted ANOVA is employed (Table 12). As before, the sample size makes non-normality a non-issue for the ANOVA.

Response: Visuality

	Sum Sq	Df	F value	p-value
Treatment	51.18	1	70.8741	0.0000
Gender	2.24	1	3.0991	0.0789
Residuals	431.11	597		

Table 12. Weighted ANOVA using treatment and gender as factors

Similarly, I ran a test comparing the range or spread of answers given on derivative and antiderivative questions by combining graphical and algebraic derivative questions into one variable, combining graphical and algebraic antiderivative questions into another variable, and taking into account gender as a factor (Table 13). Multiple previous analyses had shown that the interaction factor between gender and treatment was not significant, and thus I did not include it in the current analyses.

Response: Visuality

	Sum Sq	Df	F value	p-value
Treatment	0.61	1	1.0517	0.3055
Gender	1.27	1	2.1829	0.1401
Residuals	346.70	597		

Table 13. ANOVA using Gender and Treatment as factors

The ANOVA shows that when considering graphical and algebraic treatments neither treatment nor genders are significant factors. I decided to check the data for non-normality and inconstant variance. I ran the Shapiro-Wilks test for normality and the test statistic was 0.8813 and the p-value was 0.0000. The data is not normal, but the sample size is large; thus, I continued and conducted Bartlett's Test. The Bartlett test statistic was 1.013 and the p-value was 0.7981. I ran the follow-up Levene's Test, and found that the results agreed with Bartlett's Test. The data has constant variance. The test statistic for Levene's Test was 0.9804 with a p-value of 0.4015. As the variances are constant, there is no need for a weighted ANOVA.

In summary, when pooling treatment data, I find that there is a difference in the visuality scores of students on graphical questions versus algebraic questions. There is not a significant difference in the visuality scores of students in regards to derivative and antiderivative questions. In each analysis, gender remains an immaterial factor.

An additional observation was that the data values of the male students seemed to have less variance, propelling another analysis beyond the original study questions put forth in the introduction: what is the difference? This can be seen in Figure 4.

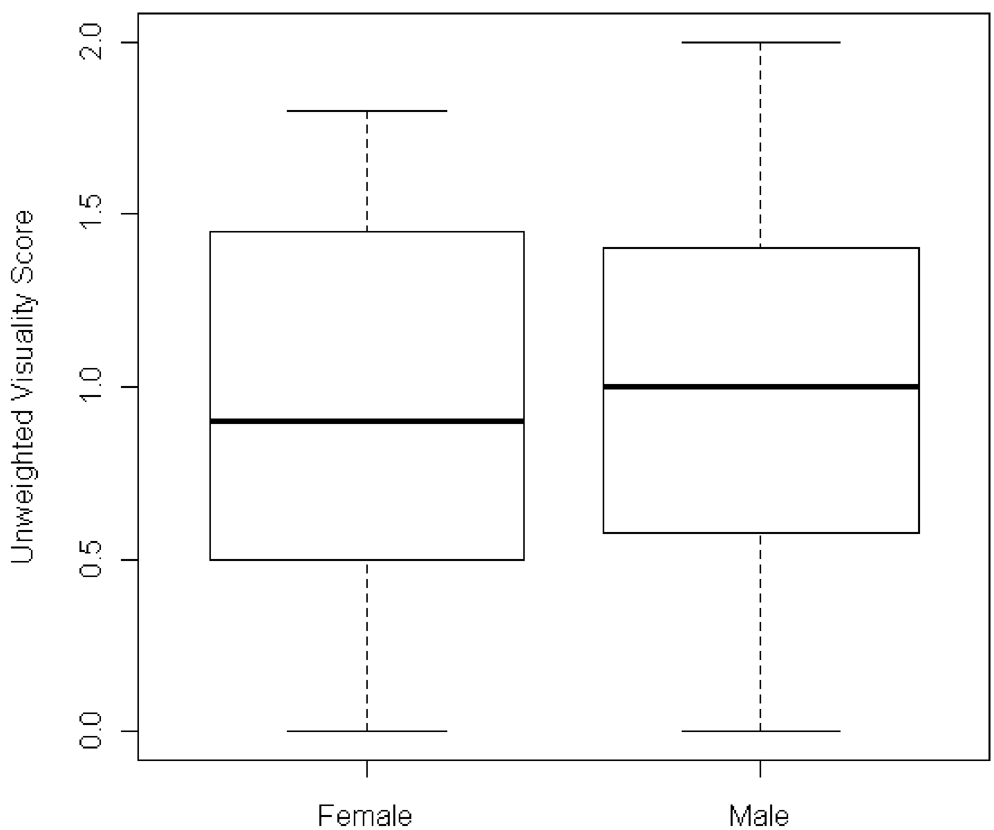

Figure 4. Box-and-Whisker plot of the unweighted visuality scores for males and females

The male students had a larger difference in their visualities for algebraic and graphical questions than female students did. Girls were more visual on graphical questions than they were on algebraic questions, but the differences for males were more extreme.

Curious as to how large the differing variances and visuality scores were between each type of question, I began testing each gender subset to see which questions were treated the same and which questions were had significantly different visuality scores. Were the results consistent between genders? I created a new variable for the difference between graphical and algebraic visuality scores and conducted an ANOVA with the difference as the response (Table 14). From the ANOVA I concluded that there is a significant difference in the response when gender is a factor.

Response: Difference between Graphical and Algebraic Visuality Scores

	Sum Sq	Df	F value	p-value
Gender	2.190	1	4.865	0.02895
Residuals	66.636	148		

Table 14. ANOVA using Gender as a factor

According to the Shapiro-Wilks test statistic, which was 0.9854, and the p-value of 0.1138, the data is normal; see Figure 5. The Bartlett test statistic was 1.1621 and the p-value for Bartlett's test was 0.2810, so I did not reject the null hypothesis of equal variances. I confirmed this conclusion with Levene's Test and found that Levene's Test agrees with Bartlett's Test. Levene's Test reported a test statistic of 0.3033 and a p-value of 0.5827.

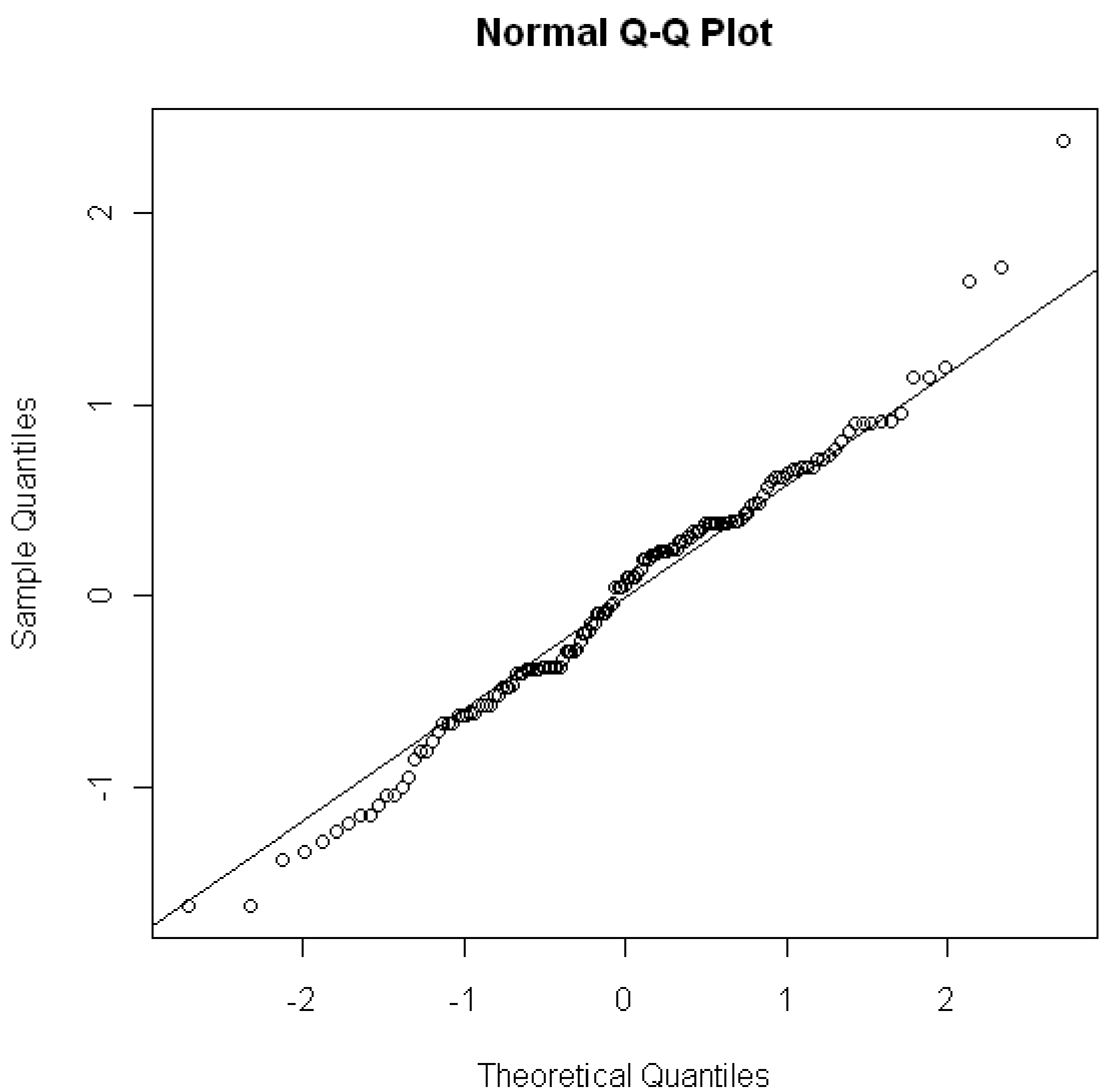

Figure 5. Q-Q Plot for Normality

Behaviors of Highly Visual, Highly Analytic Students

A separate analysis was performed on the students who exhibited the most extreme visuality scores. The top twenty five percent of students, those who were the most visual, were placed into a single subgroup. Similarly, the lowest twenty five percent of students, those who were most analytic, were also placed into another subgroup. Any students in between were categorized as harmonic. To analyze the propensity of a high analytic or high visual student changing their preference when encountering more difficult derivative and antiderivative tasks, a Chi-Square test [21] was conducted on the proportion of change in visuality score between the two extreme groups.

The p-value from the first chi-square test was 0.1862 which revealed that the proportion of students who changed their preference from analytic to harmonic was not significantly different from the proportion of students who changed from visual to harmonic. Neither was the proportion of change from analytic to visual significantly different from the proportion of change from visual to analytic. However, the assumptions for the chi-square test were not met since the expected values were not all greater than five (Table 15).

Observed Number of Types of Changes

	None	To Harmonic	To Opposite
Visual	27	7	0
Analytic	25	15	1

Expected Number of Types of Changes

	None	To Harmonic	To Opposite
Visual	23.57333	9.973333	**0.4533333**
Analytic	28.42667	12.026667	**0.5466667**

Table 15. Pearson's Chi-Square Test

To account for this, I combined all changes in preference into one group. The visuality scores were described as either having 'no change' or 'change'. The output from the altered chi-square test satisfied the assumption of each group having more than five observations. The p-value for the test was 0.2063. I concluded that the proportion of students who changed their preference from highly visual was not significantly different from the proportion of students who changed their preference from highly analytic (Table 16).

Observed Number of Types of Changes

	No Change	Change
Visual	27	8
Analytic	25	16

Expected Number of Types of Changes

	No Change	Change
Visual	23.94737	11.05263
Analytic	28.05263	12.94737

Table 16. Pearson's Chi-squared test with Yates' continuity correction

To further examine the data, both the high analytic students and the high visual students were compared to those students who were categorized as harmonic. A chi-square test was performed to analyze their proportions of change. The p-value was 0.3200 which suggested that all three groups of students shared the same proportions of change in preference when confronted with a series of more difficult antiderivative and derivative tasks (Table 17).

Observed Number of Types of Changes

	No Change	Change
Visual	27	8
Analytic	25	16
Harmonic	51	24

Expected Number of Types of Changes

	No Change	Change
Visual	23.87417	11.12583
Analytic	27.96689	13.03311
Harmonic	51.15894	23.84106

Table 17. Pearson's Chi-squared test

Conclusions

In the present study, I investigated the mediating methods of thought in mathematics; in particular, the preference for visual versus non-visual strategies of North Floridian high school students taking AP Calculus. The students' preferences for a certain method of thought were measured using the Mathematical Processing Instrument for Calculus (MPIC). The MPIC calculated their visuality score on a scale that ranged from zero to two; zero meaning their visual preference was highly analytic and two meaning their visual preference was highly visual. A visuality score of one was considered to be harmonic, meaning the student took equally from both visual and analytic methods when attempting to solve a mathematics problem. It was my goal to explore these relationships in order to better understand visual preference with regards to gender, question type, and question difficulty.

My results suggested that the students as a whole approached graphical (calculus) problems similarly, regardless of whether the problems involved derivative or antiderivative tasks. Likewise for algebraic problems, the students' visuality scores did not seem to alter when facing tasks involving derivatives and antiderivatives. The trend was not found when analyzing visuality scores associated with derivative or antiderivative questions. In short, I found that students approached these types of questions differently, depending on whether the question was graphical or algebraic in nature. None of the tests I conducted suggested that visuality scores changed significantly across gender.

Further analysis revealed, however, that females changed their method of thought more frequently than males did. While a significant change in visuality score was found when analyzing the individual differences for females as they changed from graphical antiderivative to algebraic antiderivative, as well as when they changed from graphical antiderivative to algebraic derivative questions, no such significant change in visuality score was observed amongst the male students when the individual differences where similarly analyzed.

It was not observed that highly visual or highly analytic students changed their preference statistically more or less than the other. This was also observed when comparing high visual and high analytic students to the harmonic students in the sample.

It is important that my study does not attempt to draw any conclusions that are beyond the limitations of the methods employed. A limitation of this study is that an ANOVA to compare each student to the rest of the sample was not performed.

The findings of this study are in agreement with Lowrie and Kay [5], as I saw no conclusive difference in preference for visual or analytical methods of thought across gender. The findings of this study are then contradictory to the findings of Fennema, Carpenter, Jacobs, Franke, and Levi [4], as my findings do not support the difference in visuality scores evidenced in their work.

Further studies that could come from this could seek to discover whether the ability of a student to perform mathematically can be explained by their preference for analytical, visual, or even harmonic methods of thought.

APPENDIX

Gender only ANOVA:

I analyzed how large on average the variance was between genders using visuality as the only factor. I began by creating two subsets of male and female students. I first ran an ANOVA on the unweighted mean for visuality (CV_U) with just gender as a factor (Table 18). The ANOVA shows that gender is not a significant factor.

Response: CV_U

	Sum Sq	Df	F value	p-value
Gender	0.086	1	0.2751	0.6007
Residuals	46.087	148		

Table 18. ANOVA of Visuality (Unweighted Mean)

I conducted the Shapiro-Wilks test for normality and the test statistic was 0.9596 with a p-value of 0.0002. The data is not normal, but the sample size is large. I feel comfortable proceeding with the ANOVA. I also ran Bartlett's Test, which had a statistic of 0.4933 and a p-value of 0.4825. I do not reject the null hypothesis of equal variances. To confirm that Bartlett's Test is not failing due to its sensitivity to non-normal data, I also ran Levene's Test. The test statistic for Levene's Test was 0.6816 with a p-value of 0.4104. The output from this test agrees with Bartlett's Test, so I feel comfortable proceeding with the unweighted ANOVA.

I conducted a nonparametric Kruskal-Wallis rank sum test on the unweighted mean for visuality scores. The test statistic was 0.2029 and the p-value was 0.6524. Gender is not a significant factor. I ran a nonparametric ANOVA using gender as a factor (Table 19). It agrees with both the parametric ANOVA and the Kruskal-Wallis test that gender is not a factor

	Df	SumsOfSqs	MeanSqs	F.Model	p-value
Gender	1	0.086	0.085655	0.27507	0.609
Residuals	148	46.087	0.311398	0.99814	
Total	149	46.173	1.00000		

Table 19. Nonparametric ANOVA on Unweighted Mean for Visuality Scores

Similarly, I ran an ANOVA on the weighted mean visuality score using gender again as the only factor (Table 20). The ANOVA shows that gender is an insignificant factor on visuality.

Response: CV_W

	Sum Sq	Df	F value	p-value
Gender	0.347	1	1.3081	0.2546
Residuals	39.206	148		

Table 20. ANOVA of Visuality (Weighted Mean)

In addition to the parametric ANOVA, I conducted the Shapiro-Wilks test to check the assumption of normality for the ANOVA. The statistic for the Shapiro-Wilks test was 0.9767 with a p-value of 0.0119. The data is not normal; however, I proceeded with the ANOVA because of the large sample size. Bartlett's test statistic was 1.2130 and the p-value was 0.2707. I did not reject the null hypothesis of equal variances. To confirm that Bartlett's test didn't fail due to non-normal data, I also conducted Levene's Test. Levene's test agrees with Bartlett's test with a test statistic of 0.8066 and a p-value of 0.3706. I am confident that the data has equal variances.

I ran the nonparametric Kruskal-Wallis rank sum test, and with a test statistic of 1.1655 and a p-value of 0.2803, it agrees with the ANOVA. Gender is not a significant factor. I conducted another nonparametric ANOVA on the weighted mean for visuality scores (Table 21). Its p-value agrees with both the parametric ANOVA and the Kruskal-Wallis test.

	Df	SumsOfSqs	MeanSqs	F.Model	p-value
Gender	1	0.347	0.34653	1.3081	0.248
Residuals	148	39.206	0.26491	0.99124	
Total	149	39.553	1.00000		

Table 21. Nonparametric ANOVA on the weighted mean for visuality scores

Non interaction ANOVA

From previous analyses, I already knew that a weighted ANOVA must be used on the data. I conducted the Shapiro-Wilks test and found that the data was not normal as the test statistic was 0.9481 and the p-value was 0.0000. The sample size is large, however, so I continued with the weighted ANOVA using the factors treatment and gender without the interaction factor. The factor gender is not significant; treatment, however, is significant with a p-value of 0.0000 (Table 22).

	Sum Sq	Df	F value	p-value
Treatment	52.92	3	24.4751	0.0000
Gender	2.26	1	3.1398	0.0769
Residuals	428.80	595		

Table 22. Weighted ANOVA using Treatment and Gender as factors

Kruskal-Wallis Multiple Comparison Tests with interaction term

Treatment	Value
(DGV, Female)	1
(DGV, Male)	2
(DAV, Female)	3
(DAV, Male)	4
(AGV, Female)	5
(AGV, Male)	6
(AAV, Female)	7
(AAV, Male)	8

Table23a: Reference Table for Multiple Comparison Test after Kruskal-Wallis

Comparisons			
1-2	4.86890847	88.93353	FALSE
1-3	152.94776119	93.55650	TRUE
1-4	94.49541449	88.93353	TRUE
1-5	24.38805970	93.55650	FALSE
1-6	33.79059522	88.93353	FALSE
1-7	152.91791045	93.55650	TRUE
1-8	109.94722172	88.93353	TRUE
2-3	148.07885272	88.93353	TRUE
2-4	89.62650602	84.05670	TRUE
2-5	19.51915123	88.93353	FALSE
2-6	28.92168675	84.05670	FALSE
2-7	148.04900198	88.93353	TRUE
2-8	105.07831325	84.05670	TRUE
3-4	58.45234670	88.93353	FALSE
3-5	128.55970149	93.55650	TRUE
3-6	119.15716598	88.93353	TRUE
3-7	0.02985075	93.55650	FALSE
3-8	43.00053947	88.93353	FALSE
4-5	70.10735479	88.93353	FALSE
4-6	60.70481928	84.05670	FALSE
4-7	58.42249595	88.93353	FALSE
4-8	15.45180723	84.05670	FALSE
5-6	9.40253552	88.93353	FALSE
5-7	128.52985075	93.55650	TRUE
5-8	85.55916202	88.93353	FALSE
6-7	119.12731523	88.93353	TRUE
6-8	76.15662651	84.05670	FALSE
7-8	42.97068873	88.93353	FALSE

Table 23b. Multiple comparison test after Kruskal-Wallis (p.value: 0.05)

Pairwise Comparison Test

Treatment	Value	Gender	Value
DGV	1	Female	1
DAV	2	Male	2
AGV	3		
AAV	4		

Table 24a: Reference Table for Pairwise Comparisons

Treatment:Gender

	diff	lwr	upr	p adj
2:1-1:1	-0.69154229	-1.0706524	-0.31243221	0.0000012
3:1-1:1	-0.11513859	-0.4942487	0.26397148	0.9837228
4:1-1:1	-0.67164179	-1.0507519	-0.29253171	0.0000028
1:2-1:1	-0.03164898	-0.3920259	0.32872789	0.9999952
2:2-1:1	-0.43727147	-0.7976483	-0.07689460	0.0059353
3:2-1:1	-0.13319804	-0.4935749	0.22717884	0.9514119
4:2-1:1	-0.48948031	-0.8498572	-0.12910343	0.0010688
3:1-2:1	0.57640370	0.1972936	0.95551377	0.0001243
4:1-2:1	0.01990050	-0.3592096	0.39901058	0.9999999
1:2-2:1	0.65989330	0.2995164	1.02027018	0.0000011
2:2-2:1	0.25427081	-0.1061061	0.61464769	0.3864542
3:2-2:1	0.55834425	0.1979674	0.91872113	0.0000827
4:2-2:1	0.20206198	-0.1583149	0.56243885	0.6837184
4:1-3:1	-0.55650320	-0.9356133	-0.17739312	0.0002560
1:2-3:1	0.08348961	-0.2768873	0.44386648	0.9968554
2:2-3:1	-0.32213288	-0.6825098	0.03824399	0.1188740
3:2-3:1	-0.01805944	-0.3784363	0.34231743	0.9999999
4:2-3:1	-0.37434172	-0.7347186	-0.01396484	0.0352458
1:2-4:1	0.63999281	0.2796159	1.00036968	0.0000027
2:2-4:1	0.23437032	-0.1260066	0.59474719	0.4975500
3:2-4:1	0.53844375	0.1780669	0.89882063	0.0001791
4:2-4:1	0.18216148	-0.1782154	0.54253836	0.7867815
2:2-1:2	-0.40562249	-0.7462374	-0.06500756	0.0076046
3:2-1:2	-0.10154905	-0.4421640	0.23906588	0.9853900
4:2-1:2	-0.45783133	-0.7984463	-0.11721639	0.0012743
3:2-2:2	0.30407344	-0.0365415	0.64468837	0.1199114
4:2-2:2	-0.05220884	-0.3928238	0.28840610	0.9997867
4:2-3:2	-0.35628227	-0.6968972	-0.01566734	0.0329517

Table 24b. Pairwise Comparison Test

References

[1] Haciomeroglu, E. S., Aspinwall, L., Presmeg, N., Chicken, E., & Bu, L. (2009). Mathematical Processing Instrument for Calculus. *ON-Math: Online Journal of School Mathematics, 7*(1).

[2] Carr, M. (2008) A Comparison of Predictors of Early Emerging Gender Differences in Mathematics Competency. *Learning and individual differences,* 18:61-75.

[3] Orhun, N. (2007) An Investigation into the Mathematics Achievement and Attitude Towards Mathematics with Respect to Learning Style According to Gender. *International Journal of Mathematical Education in Science and Technology* 38.3:321-333.

[4] Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L.W., A Longitudinal Study of Gender Differences in Young Children's Mathematical Thinking, *Educ. Res.* 27(1998), pp. 6–11.

[5] Lowrie, T., & Kay, R. (2001). Relationship between Visual and Nonvisual Solution Methods and Difficulty in Elementary Mathematics. *The Journal of Education Research, 94*(4), 248-255.

[6] Krutetskii, V. A. (1976). *The Psychology of Mathematical Abilities in Schoolchildren*. Chicago: University of Chicago Press.

[7] Moses, B. E. (1977) *The Nature of Spatial Ability and its Relationship to Mathematical Problem Solving*, Ph.D. diss., Ohio State University.

[8] Suwarsono, S., (1982) *Visual Imagery in the Mathematical Thinking of Seventh Grade Students*, Ph.D. diss., Monash University.

[9] Ferrini-Mundy, J., (1987) Spatial Training for Calculus Students: Sex Differences in Achievement in Visualization Ability, *J. Res. Math. Educ.* 18: 126–140.

[10] Ubuz, B., (2007) Interpreting a Graph and Constructing its Derivative Graph: Stability and Change in Students' Conceptions, *Int. J. Math. Educ. Sci. Technol.* 38: 609-637.

[11] Chambers, J. M. and Hastie, T. J. (1992) *Statistical Models in S*, Wadsworth & Brooks/Cole.

[12] Siegel and Castellan (1988) Non-parametric Statistics for the Behavioural Sciences. MacGraw Hill Int., New York. pp 213-214.

[13] Anderson, M.J., (2001) A New Method for Non-parametric Multivariate Analysis of Variance. *Austral Ecology*, 26: 32–46.

[14] McArdle, B.H., & Anderson, M.J., (2001) Fitting Multivariate Models to Community Data: A Comment on Distance-based Redundancy Analysis. *Ecology*, 82: 290–297.

[15] Patrick Royston (1982) An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Applied Statistics*, **31**, 115–124.

[16] Bartlett, M. S. (1937). Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London Series A* **160**, 268–282.

[17] Fox, J. (2008) *Applied Regression Analysis and Generalized Linear Models*, Second Edition. Sage.

[18] Myles Hollander and Douglas A. Wolfe (1973), *Nonparametric Statistical Methods.* New York: John Wiley & Sons. Pages 115–120.

[19] Salkind, N. J., (2010) *Encyclopedia of Research Design.* Sage Publications.

[20] Holm, S. (1979). "A Simple Sequentially Rejective Multiple Test Procedure". *Scandinavian Journal of Statistics* **6** (2), 65–70.

[21] Hope, A. C. A. (1968) A Simplified Monte Carlo Significance Test Procedure. *J. Roy, Statist. Soc. B* **30**, 582–598.