Factors Related to the Increased Percentage of Public High School Students Attending College in the Northeastern United States

Code: DA-006

Executive Summary

As a result of the ever-present pressure placed upon public school administrators to make cost-effective economic and administrative decisions, a study has been conducted with the goal of assessing student achievement in comparison to a variety of school-wide statistics. The study was conducted using data from a set of 135 public high schools in the Northeastern United States. Student achievement was measured according to the percentage of students attending college. As expected, the percentage of students attending college was shown to be highly related to the mean performance on standardized tests. After examining the percentage of students attending college for all schools within the study, it was evident that schools with a lower average dropout rate and a higher average SAT participation rate witnessed a higher average percentage of students going to college. Both of these factors may imply that improving student motivation and interest in taking the steps necessary to attend college should be the focus of future school initiatives. Contrary to the common belief that high cost per pupil and low student to teacher ratios result in greater student achievement, the study failed to find any significant relationship between these factors and college attendance. In addition, schools with a higher average teacher salary showed a minor tendency to have higher proportions of students attending college and schools with a higher average enrollment showed a minor tendency to have increased average test scores, which in turn increased the percentage of college attendees.

Introduction

Throughout the years, public schools frequently face the challenges of maintaining a set of high educational standards, while working with a somewhat dynamic budget. As government funding fluctuates in accordance with policy changes and state fiscal allowances, public high schools and school districts are often forced to make administrative and budgetary decisions with the intention of maximizing the benefits that the student population will receive. In order to determine specific factors with which to focus monetary and personnel resources, a study was sponsored to assess student achievement. Information was gathered and statistics computed for 135 public high schools in the Northeastern United States. To evaluate student achievement, it was decided that the percentage of students going to college would be the variable of interest. In addition to college attendance, standardized test scores are frequently used as an indicator of high school student achievement. In this case, however, it was decided to focus mainly on college percentages. This assumption was made as a result of the high observed correlation between test scores and college attendance in addition to the fact that scores on tests such as the SAT, are known to directly influence college admission. Information was also provided concerning the percentage of students per school attending one of four different types of colleges (public two-year, private two-year, public four-year, and private four-year). Although this data was analyzed, it was not used to draw conclusions about overall college attendance. After examining the relationship between type of college and cost-related variables, it was evident that schools with higher average teacher salaries and higher average cost per pupil were related to a higher percentage of private and four-year college attendance. Under the assumption that these variables may indicate higher income neighborhoods, it did not seem logical to measure student achievement based on a specific college type, when family income may be an important factor in the decision students make to attend a specific type of college. For figures related to this conclusion please see Appendix A.

The main concern of this analysis will be to determine which economic and administrative factors should be the focus of future initiatives aimed at increasing the percentage of high school students going to college. First, hypothesis testing will be performed concerning the comparison of two groups. Groups will be created based on schools with a high percentage of college attendance and schools with a low percentage. T-tests and nonparametric tests can then be conducted to test the hypotheses that the mean values of specific variables differ significantly between two distinct college attendance categories. Next, the college attendance variable will be categorized into three separate groups and a one-way analysis of variance (ANOVA) will be performed to better understand the differences that set schools with the highest percentage of college attendance apart. As expected, while conducting this analysis, the hypothesis that increased average test scores on both the MCAS (Massachusetts Comprehensive Assessment System) and the SAT (Verbal and Scholastic Assessment) correspond to an increased percentage of students attending college will be accepted. Based on the acceptance of this hypothesis and the evaluation of variable correlations, two multiple regression models will be built to determine which variables are highly related to both college attendance and test scores. After completing these analyses, strong conclusions can be drawn regarding the factors that are expected to increase college attendance rates, both directly and through increased test scores.

Summary Information

Before completing any advanced analysis, a look into the summary statistics will help to better describe and understand the dataset structure. The dataset includes the information from 135 schools with no missing values. The mean and standard deviation of each variable included in the study are shown in Figure 1.

Descriptive Statistics

	Mean	Standard Deviation
Enrollment	1,103.39	572.43
Cost per pupil	7,111.96	1,192.01
Average Teacher Salary	46,963.23	4,468.55
Average SAT Verbal Score	512.64	41.83
Average SAT Math Score	518.25	44.26
SAT Participation Rate	0.81	0.13
Average MCAS English Score	243.42	7.11
Average MCAS Math Score	241.83	7.48
Student/Teacher Ratio	13.13	1.94
Student/Counselor Ratio	224.52	53.00
Dropout Rate	2.05	2.21
Percentage Going to College	82.92	7.87
Percentage Going to 2yr Public Colleges	13.41	8.69
Percentage Going to 4yr Public Colleges	27.05	7.79
Percentage Going to 2yr Private Colleges	3.34	4.36
Percentage Goint to 4yr Private Colleges	39.12	14.64
Percentage Going into the Military	2.10	1.72
Percentage Going to Work	9.21	6.02
Percentage Choosing Other Activities	5.80	5.03

Figure 1

It is important to note, that of the variables related to high school conditions, both dropout rate and enrollment have fairly large standard deviations in relation to their mean values. This indicates that these particular variables may be increased or decreased within individual schools with the possibility of positively affecting college attendance. Therefore, these variables display flexibility. Figure 2 displays the distribution of post-graduation pursuits on average for all 135 schools. Within the sample, an average of approximately 83% of students are recorded as attending college after graduation. The objective of this report will be to increase this overall percentage by isolating factors that can raise individual school percentages to a level above this overall average. From examining Figure 2, it is evident that the percentage of students going to college is currently fairly high, in comparison to the percentage of students going to the military, going to work and choosing other activities. Since the distribution in Figure 2 will necessarily have to be shifted in order to increase college attendance, it is reasonable to assume that there will be some maximum percentage above which it will be difficult to raise the college attendance percentage. For example, despite all efforts at the school level a certain percentage of students will still be expected to enroll in the military, out

of personal preference. However, as Figure 1 shows, the standard deviation of college attendance is close to 8%. Therefore, it is not unreasonable to maintain an objective of moderately raising this percentage for the set of schools with current college attendance rates at the average or below average level.

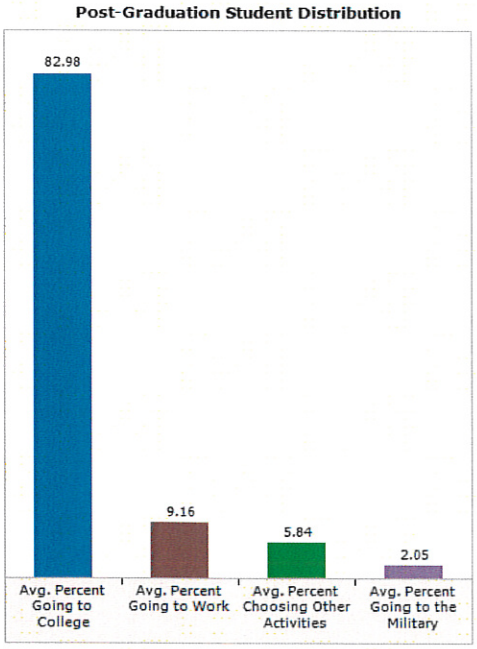


Figure 2

Statistical Analyses

To begin the analyses, it was decided to focus on differences that existed between schools that could be classified as having high college attendance rates and those that could be classified as having relatively low college attendance rates. To do this the continuous variable, percentage of students going to college was categorized. Schools were separated on the basis of an 84% college attendance rate cut-off, in which schools with less than 84% of their students attending college were placed in one group and those with greater than or equal to 84% were placed in another group. This decision was made in order to create as balanced a design as possible. With the goal in mind of testing hypotheses related to the comparison of various mean values, it was important to determine whether the theoretical distribution of sampling means for the variables followed a normal distribution. With a sample size of 135, it was important that the sample values fell into a somewhat normal distribution. Through the analysis of qqplots and histograms it was determined that the majority of the variables of interest could be classified as normal, with a few exceptions. Both variables enrollment and dropout rate were highly right-skewed and it was possible to successfully perform a log transformation on these variables to create normal distributions (all values of zero for dropout rate were given a dummy vale of 0.5 to ensure that these variables would not be dropped from the dataset). Details of the normality assumption checking, including all figures may be found in Appendix B. For all values that passed a Kolmogorov-Smirnov test for normality, a t-test was performed and the results are displayed in Figure 3. Based on the assumptions of the t-test, it was necessary to test for equal group variances. For each variable, the probability of the variances being unequal due to chance was computed. For variables with probability greater than 0.05, the equal variances t-test was performed using pooled estimation. For variables with probability less than 0.05, the unequal

Comparison of Means

				Percentage of Students Attending College (Categories)			
				High College Percentage		Low College Percentage	
Variable	P-value	Test	Variance	Mean	Std Dev	Mean	Std Dev
Enrollment	0.0072	T-Test	Equal	957.20	380.91	1,243.20	682.98
Cost per Pupil	0.0704	K-W	Equal	7,257.20	1,154.30	6,973.10	1,219.10
Average Teacher Salary	0.0139	T-Test	Equal	47,925.00	4,883.80	46,043.00	3,846.00
Average SAT Verbal Score	<.0001	T-Test	Equal	535.86	31.69	490.42	38.26
Average SAT Math Score	<.0001	T-Test	Equal	543.77	38.94	493.84	34.26
SAT Participation Rate	<.0001	K-W	Unequal	0.89	0.08	0.72	0.12
Average MCAS English Score	<.0001	T-Test	Unequal	247.55	4.91	239.48	6.67
Average MCAS Math Score	<.0001	T-Test	Equal	246.52	5.32	237.35	6.45
Student/Teacher Ratio	0.0860	K-W	Equal	12.82	1.69	13.44	2.12
Student/Counselor Ratio	0.2254	T-Test	Equal	218.85	52.70	229.94	53.09
Dropout Rate	<.0001	T-Test	Equal	1.11	1.19	3.02	2.52

Figure 3

variances t-test was performed using Satterthwaite's estimation. Variables that did not pass the Kolmogorov-Smirnov test for normality (at the 0.05 level) were tested using the non-parametric Kruskal-Wallis test (K-W), which was chosen for its effectiveness at testing differences in location between parallel groups (*Fleiss, 1999*). For each variable of interest, Figure 3 displays the specific test used along with the comparative means and the probabilities that these means differ by chance alone (p-values). P-values for enrollment and dropout rate variables were calculated using log transformations.

Figure 3 compares mean values for a series of school-wide variables for schools with a high percentage of students attending college and those with a low percentage of students attending college. As expected, average SAT Math and SAT Verbal scores were found to be significantly higher among schools within the high college percentage category. Also not surprisingly, schools with a higher rate of college attendance showed significantly higher mean MCAS English and MCAS Math scores. Two other factors, school dropout rate and SAT participation, were found to be highly significant (at the < 0.0001 level). SAT participation at schools with high college attendance rates averaged 89%, while participation at schools with low college attendance rates averaged only 72%. High school dropout rate shows an increased average of 3.02% for schools with low levels of college attendance as compared to only 1.11% on average for schools with high college percentages. Decreased enrollment numbers proved to be statistically significant at the 0.0072 level in schools with relatively high college attendance. In addition, increased average teacher salary was found to be significant at the 0.0139 level in schools with high percentages of students attending college.

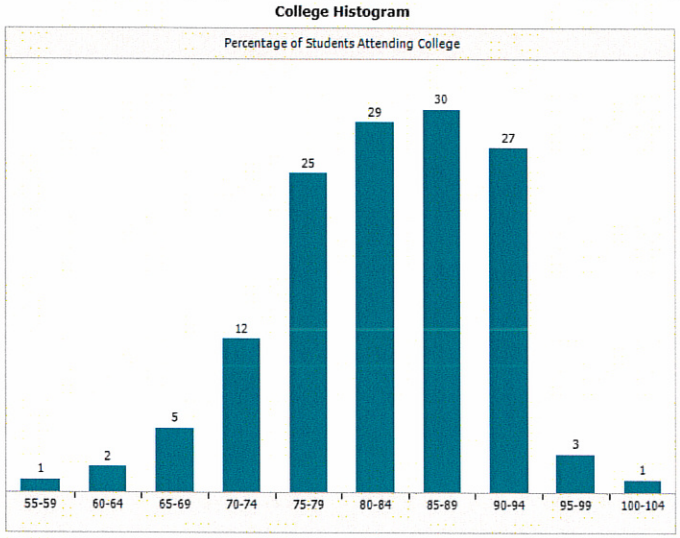


Figure 4

Figure 4 is a histogram for the college percentage variable. When examining this diagram it is evident that the majority of school percentages tend to be densely clustered around the mean of 83%. It is worthwhile to consider whether the two separate categories of college percentages (high and low) are adequately differentiating between attendance levels. Intuitively, the conclusions will be somewhat limited if the majority of the records within both the high and low college attendance categories are clustered around the same cutoff value. In order to increase the power of the comparisons the college percentage variable was divided into a set of three independent groups and a one-way ANOVA was performed to determine possible differences between means. Once again, categorization was carried out with the intention of creating a balanced design. Low college percentage is defined as schools with less than 80% of students attending college, medium college percentage is defined as schools with levels of attendance above 80%, but less than 88%, and high college percentage is defined as schools with 88% to 100% of students going to college. Although one-way ANOVA is considered a robust procedure, the same assumptions as discussed when performing t-tests should be upheld, therefore log transformations were used for variables enrollment and dropout rate. Results are shown in Figure 5.

One-Way ANOVA Results

Variable	P-Value	F-Value	College Category		
			Low College Percentage	Medium College Percentage	High College Percentage
Enrollment	0.0009	7.35	1334.76	1069.61	902.068
Cost per Pupil	0.1724	1.78	6969.2	6986.9	7388.7
Average Teacher Salary	0.0855	2.51	45819.4	47216.3	47868.5
SAT Participation Rate	<.0001	54.94	0.69022	0.82696	0.90818
Student/Teacher Ratio	0.1437	1.97	13.5111	13.1739	12.7045
Dropout Rate	<.0001	37.11	3.5822	1.7326	0.8090

Figure 5

Figure 5 compares mean values of a set of variables between low, medium and high college percentage categories. Test scores were not included in this specific analysis due to the strength of the relationship between test scores and college attendance and the intuitive nature of this relationship. Differences in enrollment, SAT participation rate, and dropout rate were all found to be statistically significant at the 95% level. In this case, the p-values indicate that differences in mean values in one or more of the three possible categories. In order to determine the specific differences between categories, a Student-Newman-Keuls' (SNK) multiple range test for post-hoc comparisons was performed (*Cody, 2006*).

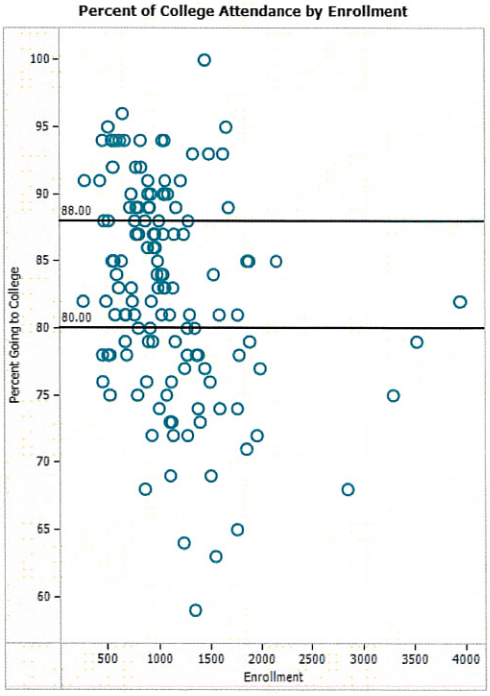


Figure 6

Schools with high percentages of students going to college had an average dropout rate of only 0.81% as compared to 3.58% for schools with low rates of college attendance. Schools with a high percentage of students going to college also had an average SAT participation rate of around 91% as compared to an average of only 69% for schools with low rates of college attendance. For both dropout rate and SAT participation the SNK test indicates that significant differences exist between all three levels of college attendance. Mean enrollment differed from 902 for schools with a high percentage of students going to college to 1,335 for schools with a low percentage of students going to college. For enrollment, the SNK test found that schools with low college percentages were significantly different from schools with medium or high college percentages. This would imply that schools with high enrollment levels may also have low college attendance rates, however once enrollment has been decreased to around 1,000 students, further decreasing enrollment would not be expected to additionally raise college attendance levels. This result is clearly shown in the plot of enrollment against college attendance in Figure 6. Reference lines have been added to indicate the three college categories of discussion. As indicated in Figure 6, schools with enrollment of 1,000 students or above tend to have lower overall percentages of college attendance. The variable for average teacher salary was found significant when a t-test was completed, however it was not found to be significant when evaluated using ANOVA. This is likely a result of the decreased sample size per group, which is a notable consequence of creating additional categories.

Next, models were designed to estimate the exact effect that certain continuous variables had on the percentage of students going to college. A multiple linear regression was performed with a subset of the variables included in the model-building steps. In the first model, all test

score statistics were not included in order to be able to focus on factors that could possibly be classified as causes, rather than indicators, of student achievement. The first model studies the dependence of the college attendance variable on the chosen subset of variables. In the second model this same subset of variables will be analyzed in relation to average MCAS score (the average of English and Math scores), with the expectation that the results will be somewhat similar to those of the first model. This expectation is due to the high correlation between MCAS score and college attendance. MCAS scores were chosen over SAT scores due to the fact that both English and Math MCAS scores showed higher positive correlation with college attendance than did either SAT Math or SAT Verbal scores (see Appendix C, Table 2). The first step in model building was to examine correlations between variables. Correlation between variables may require exclusion of a member of any set of multiple variables with high correlations to avoid creating a model with confounding independent variables. The correlation matrix used can be found in Appendix C, Table 1. Once again, log transformations were used for enrollment and dropout rate. Models were built based on comparisons of R-square values and Mallow's Cp statistics for various combinations of possible variables. After this preliminary evaluation, a forward multiple regression was conducted and a model was selected based upon high R-square values in combination with favorable Cp statistics (*Cody, 2006*). Models selected were then tested for multicollinearity using the variance inflation factor. The results of both regressions are shown in Figure 7.

Multiple Linear Regressions

Dependent Variable	R-Square	Cp	Independent Variable	P-Value	Slope Estimate	Standard Error	Variance Inflation
College Percentage	0.5782	1.6707	SAT Participation Rate	<.0001	36.6296	4.3734	1.7674
			Dropout Rate	0.003	-1.6785	0.5541	1.6816
			Student Teacher Ratio	0.1565	0.3401	0.2386	1.1190
MCAS Score	0.7301	3.2794	SAT Participation Rate	<.0001	29.8965	3.1859	1.8177
			Dropout Rate	<.0001	-2.3880	0.4052	1.8280
			Enrollment	0.1558	-1.0896	0.7632	1.4767

Figure 7

The results in Figure 7 can be used to estimate increases in college attendance. For every 10% increase in SAT participation, a 3.66% increase in the number of students going to college would be expected, if all other variables are held constant (SAT participation is expressed as a rate and college attendance is expressed as a percentage). Since dropout rate was regressed using the log of dropout rate, a decrease of 1 in the natural log of dropout rate would be expected to increase the percentage of students going to college by 1.68%. For example, if the dropout rate was decreased from 3.00% (log 3.00 is equivalent to 1.1) to 1.11% (log 1.11 is equivalent to 0.1), the percentage of students going to college would be expected to increase by 1.68%, if all other variables are held constant. For every increase of 1 in the student teacher ratio, an increase of 0.34% in the percentage of students going to college would be expected, though this estimate is not significant at the 0.05 level. According to the regression model, with a series of alterations in SAT participation rate, dropout rate, and the student/teacher ratio, it is possible to raise college attendance rates a moderate amount, which is the study objective.

The second model examines possible factors affecting the average MCAS score. For every 10% increase in SAT participation, an increase of 3 points in the average MCAS score would be expected, all other variables being held constant. A decrease of 1 in the natural log of dropout rate would be expected to decrease the average MCAS score by 2.39 points. Also, a decrease of 1 in the natural log of enrollment would be expected to raise the average MCAS score by 1.09 points, although this parameter estimate is not significant at the 0.05 level. Parameter estimates that are not significant were included in both models due to the fact that they helped

explain a greater percentage of model variance, and showed relatively low correlation to other significant variables in the model.

Conclusion

Based on the results discussed previously, it was determined that there are methods available for schools to increase student achievement, which is defined as the percentage of students going to a college of any type. The main recommendation of this analysis for schools with low or average student achievement is to promote initiatives aimed at increasing student motivation to attend college. Increased SAT participation and decreased school-wide dropout rate were two factors that were found to be strongly related to the level of college attendance. Although it may not be reasonable to assume that low SAT participation and high dropout rates cause schools to have lower percentages of students attending college, it is logical to consider that these factors are representative of a certain school-wide environment that translates to low college attendance. Financially, the allocation of funds to the promotion of SAT preparation courses, tutoring centers or support groups for struggling students may be examples of some efforts that could reduce dropout rates and increase interest in the SAT.

Aside from SAT participation and dropout rates, enrollment and average teacher salary were found to have minor affects on college attendance. Although only small differences were observed between schools with enrollments of 1,000 students or less, there was a significant decrease in college attendance for schools with enrollments above 1,000. Low levels of enrollment were also found to increase average MCAS test scores, which in turn increase percentages of students going to college. If possible, one recommendation would be to maintain school enrollments at or below 1,000 students. Mean teacher salary was shown to be significantly higher at schools with above average rates of college attendance. The difference in salary between teachers at high and low college attendance schools was close to $2,000. Perhaps salary incentives allow for the recruitment of higher quality or more experienced teachers which may in turn increase levels of college attendance.

Contrary to the common belief that low student to teacher ratios, low student to counselor ratios, and high costs per pupil cause increased student achievement, none of these variables were found to significantly increase student performance on the MCAS or the percentage of students going to college.

One of the main limitations of this study is the lack of specificity of the information provided. Data for many of the variables provided was truncated, causing the distribution of values to be lacking the ideal level of normality. Additionally, it is hard to assume that methods for increasing student college attendance at one school would necessarily be effective at another school. Information pertaining to school location, district, demographics, and average family income would be useful in creating a better analysis of differences between schools. In future studies, the analysis would be considerably more powerful if records had been collected at the student level with an individual record for each student including information such as school, specific test scores, and grade point average. This would also allow for the computation and analysis of the variance associated with each school-wide mean.

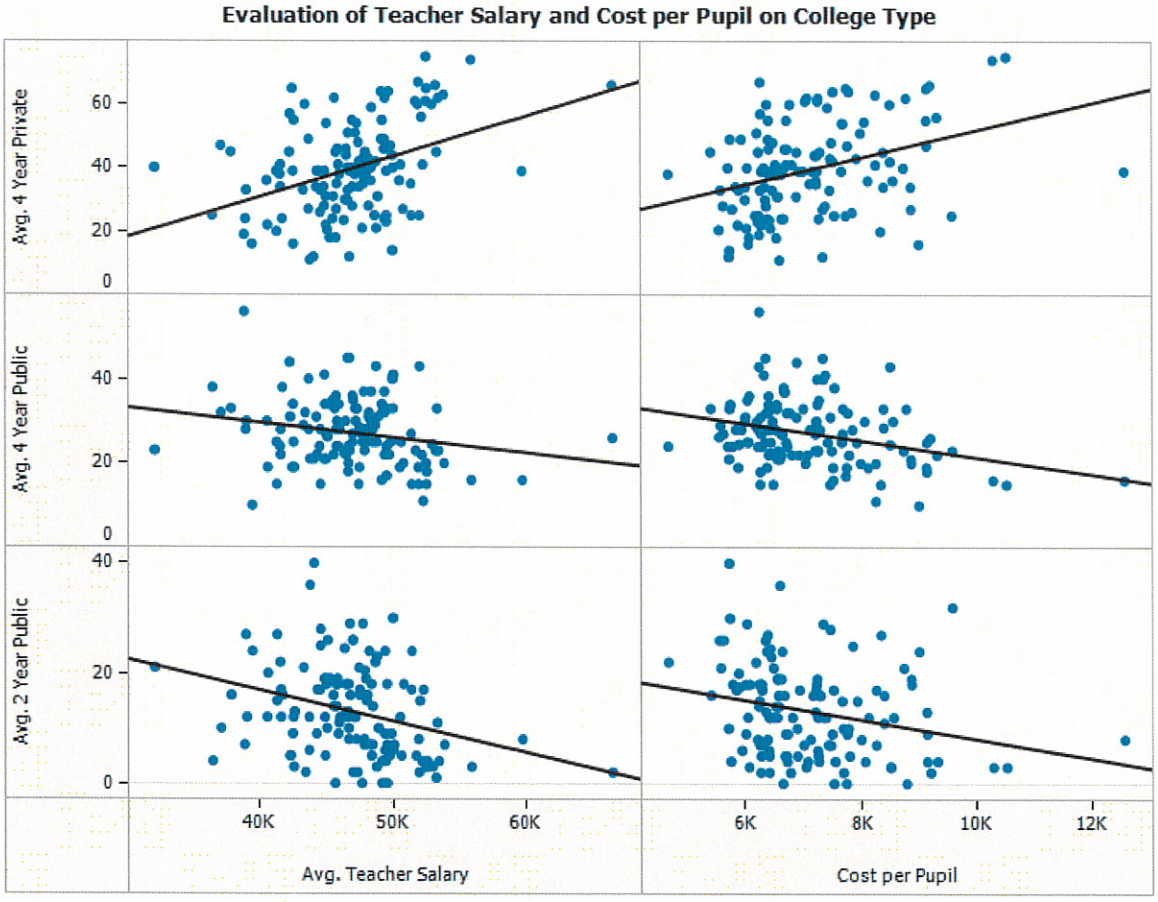
References

Cody, Ronald P., and Jeffrey K. Smith. 2006. *Applied Statistics and the SAS Programming Language,* 5th ed. Upper Saddle River, New Jersey: Pearson Prentice Hall.

Fleiss, Joseph L.. 1999. *Design and Analysis of Clinical Experiments*. New York, New York: John Wiley & Sons, Inc..

APPENDICES

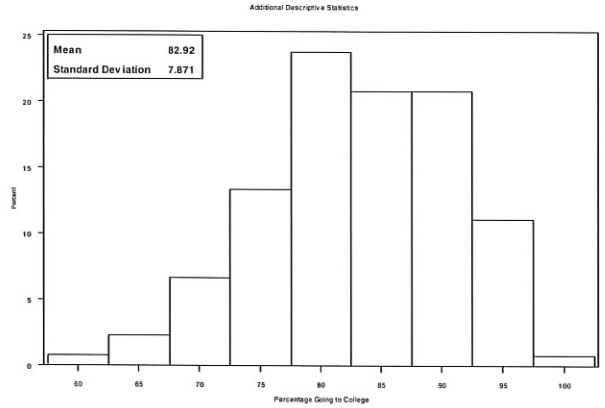
Appendix A: Evaluation of College Type Variable

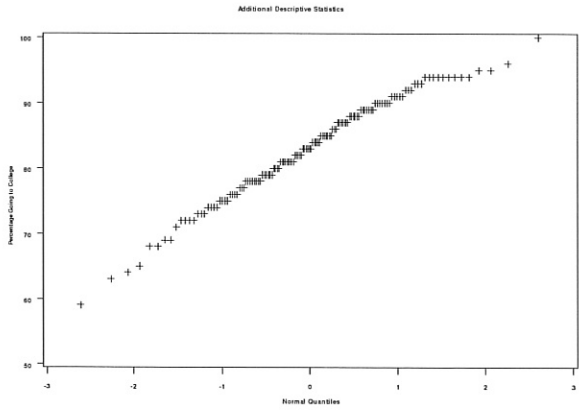


* Avg. 2 Year Private College attendance rates were not included due to low mean values.

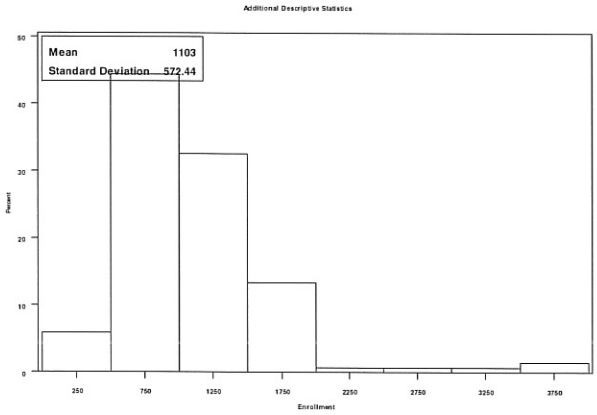
Appendix B: Normality Assumption Checking

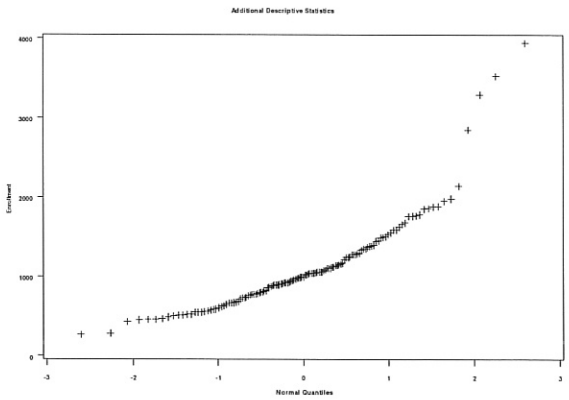
Percentage Going to College



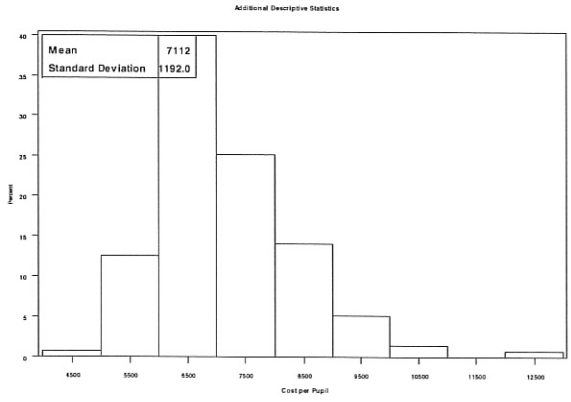


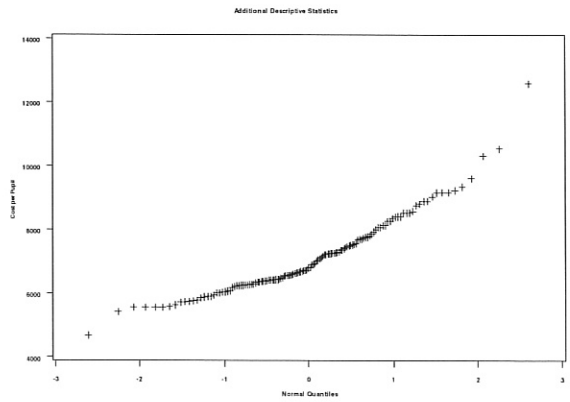
Enrollment



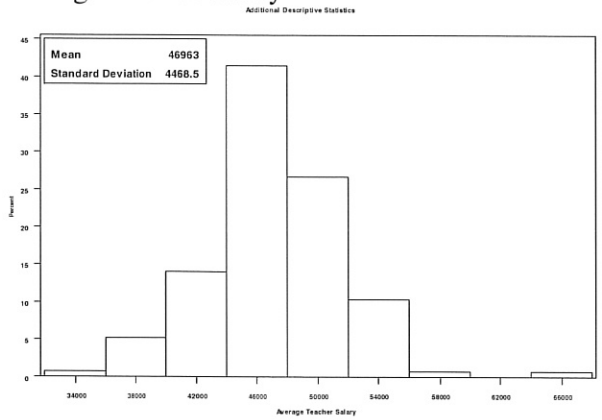


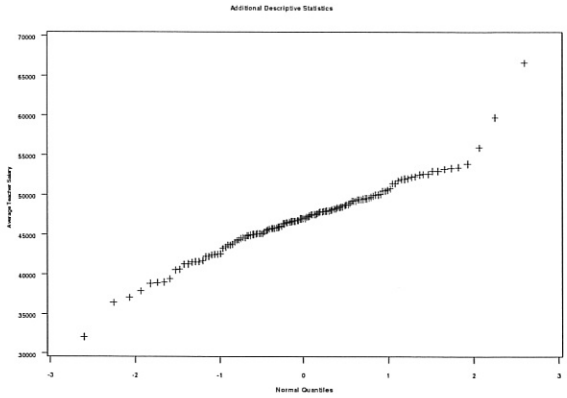
Cost per Pupil



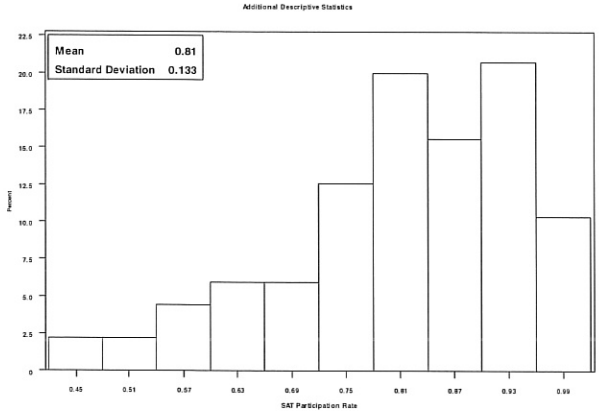


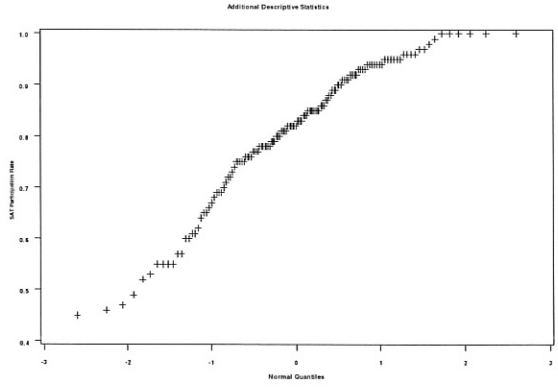
Average Teacher Salary



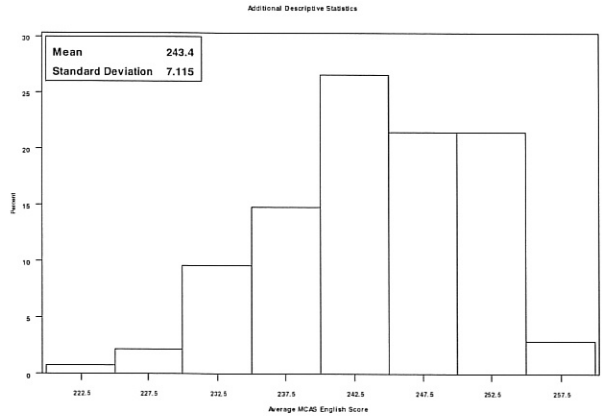


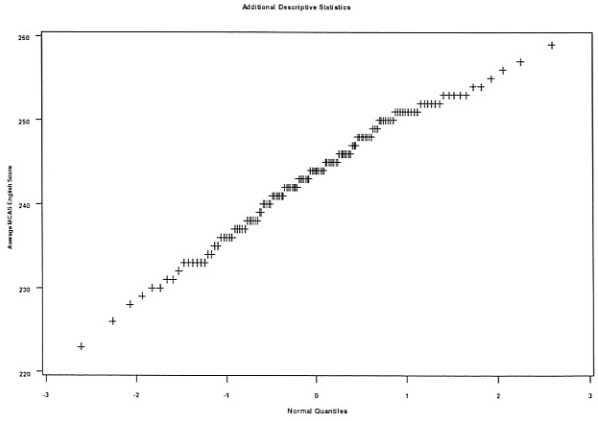
SAT Participation



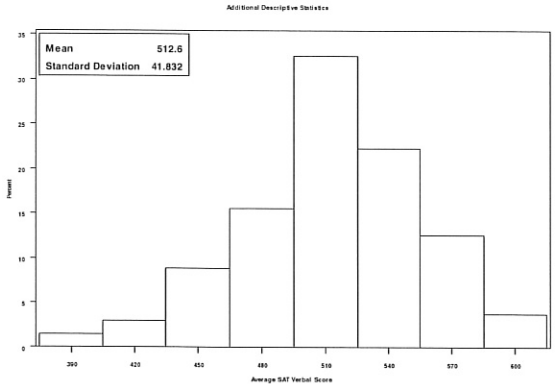


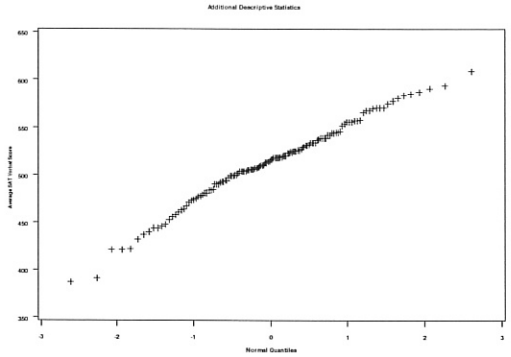
MCAS English Score



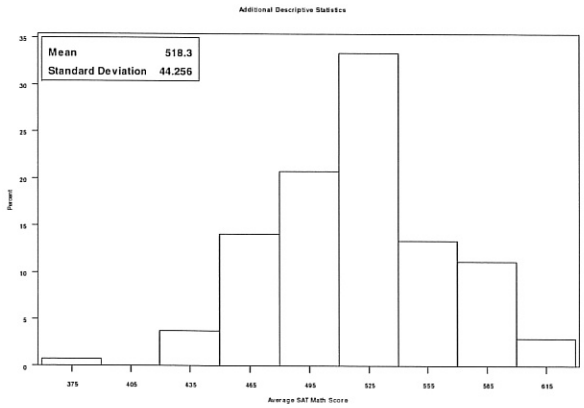


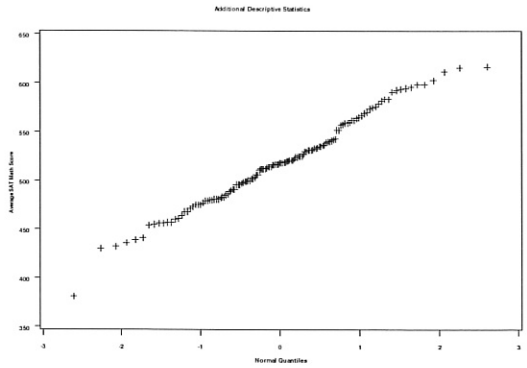
SAT Verbal



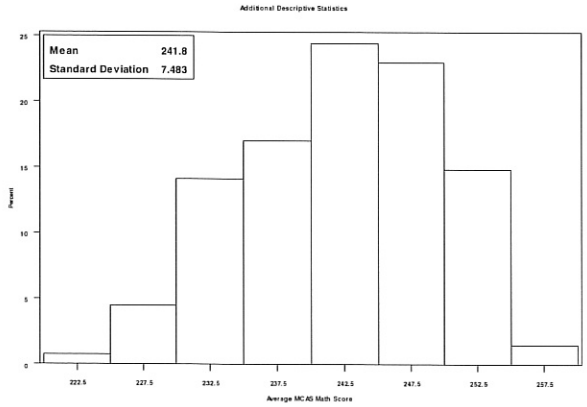


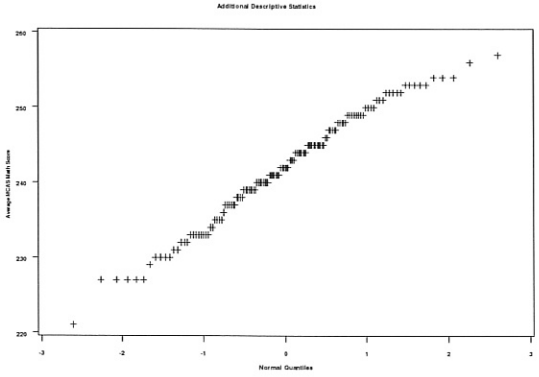
SAT Math Score



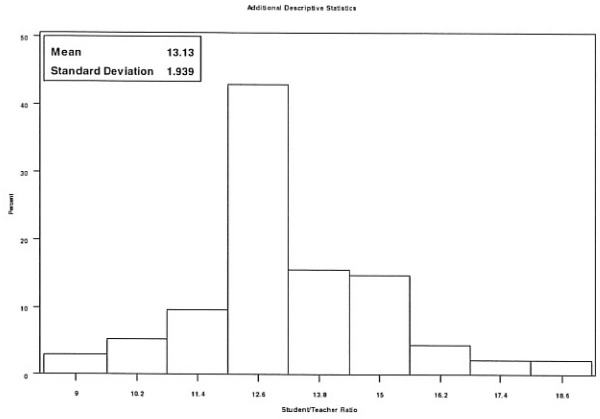


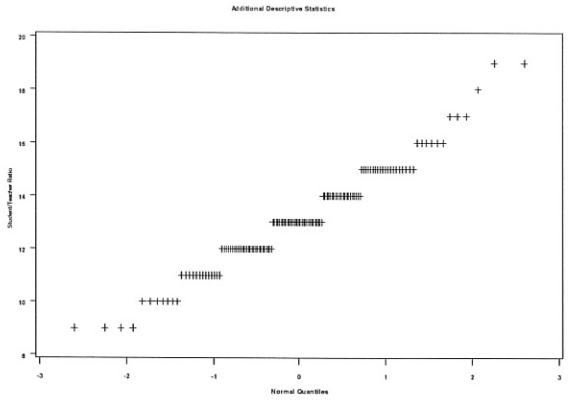
MCAS Math Score



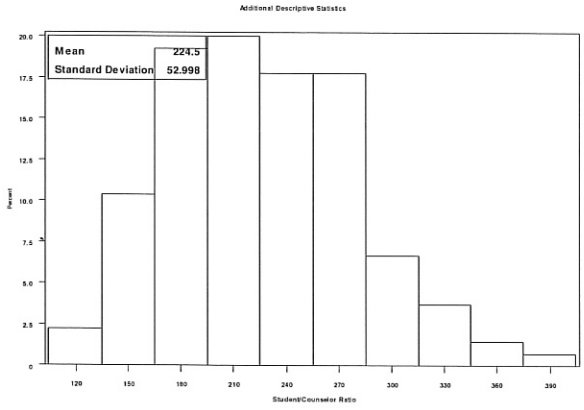


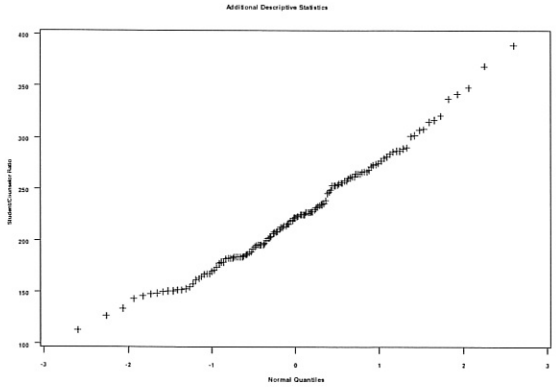
Student/Teacher Ratio



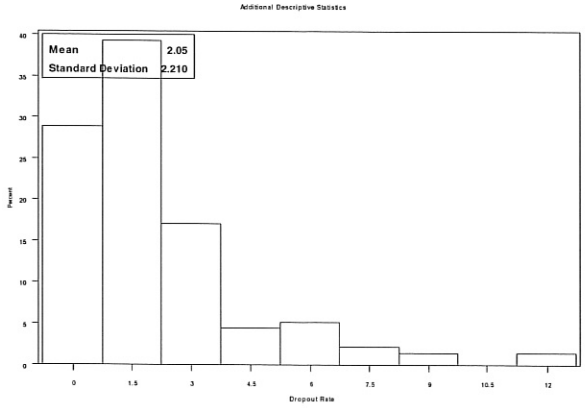


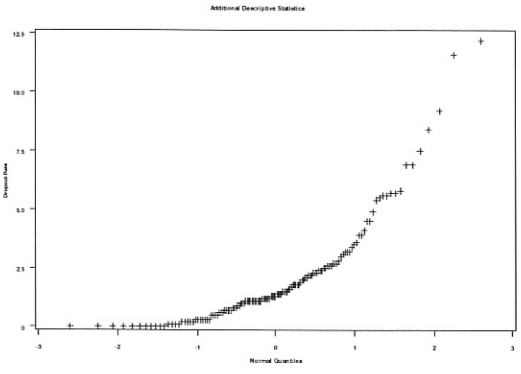
Student/Counselor Ratio





Dropout Rate





Appendix C: Correlation Matrices

Correlation Matrix

	Percentage Going to College	Enrollment	Cost per Pupil	Average Teacher Salary	SAT Participation Rate	Student Teacher Ratio	Student Counselor Ratio	Dropout Rate
Percentage Going to College	1.0000	-0.3095	0.1282	0.1937	0.7367	-0.1172	-0.0996	-0.5047
Enrollment	-0.3095	1.0000	-0.0417	0.1727	-0.4132	0.2685	0.0093	0.5391
Cost per Pupil	0.1282	-0.0417	1.0000	0.3504	0.1870	-0.4693	-0.3792	-0.1618
Average Teacher Salary	0.1937	0.1727	0.3504	1.0000	0.2466	-0.0277	-0.1454	-0.2024
SAT Participation Rate	0.7367	-0.4132	0.1870	0.2466	1.0000	-0.2593	-0.1542	-0.6185
Student Teacher Ratio	-0.1172	0.2685	-0.4693	-0.0277	-0.2593	1.0000	0.3034	0.1521
Student Counselor Ratio	-0.0996	0.0093	-0.3792	-0.1454	-0.1542	0.3034	1.0000	0.0562
Dropout Rate	-0.5047	0.5391	-0.1618	-0.2024	-0.6185	0.1521	0.0562	1.0000

Table 1

Correlation Matrix

	Percentage Going to College	Average SAT Verbal Score	Average SAT Math Score	Average MCAS English Score	Average MCAS Math Score
Percentage Going to College	1.0000	0.6259	0.5989	0.6672	0.6841
Average SAT Verbal Score	0.6259	1.0000	0.9355	0.8876	0.8958
Average SAT Math Score	0.5989	0.9355	1.0000	0.8546	0.9058
Average MCAS English Score	0.6672	0.8876	0.8546	1.0000	0.9504
Average MCAS Math Score	0.6841	0.8958	0.9058	0.9504	1.0000

Table 2