

# Hispanic Student Achievement: An Investigation of Various Factors and the Effects on Hispanic High School API scores in California

Eleni V. Papailias,  
California State University, Sacramento

The purpose of this study is to determine specific factors that influence changes in Hispanic API scores in California public high schools. The study's design examines various high school, student, and social factors that may be significantly correlated to Hispanic academic performance. The findings reveal that policy-makers and educational administrators must consider multi-faceted solutions in creating greater success in the Hispanic student population.

## I. Executive Summary

Educational achievement is a main concern for school districts, administrators, and the educational system as a whole. Recently in California, increasing attention is being paid to school performance due to a growing and diversified student population. With a public education system ranking in the bottom half among all U.S. states, California must seek to improve the system and in turn increase students' chances for success in order to benefit the state's economy by supplying a better educated workforce. However, to facilitate such improvements the focus must shift towards the booming Hispanic student population. What factors significantly affect Hispanic students' educational achievement? Perhaps such factors directly relate to various student characteristics, social dynamics, or distinctive attributes of educational intuitions. This study involves the examination of California's public high schools (N = 995) centering on 2007 base Academic Performance Index (API) scores for Hispanic high school students as to assess such dynamics.

The study connects various factors to Hispanic academic performance. A regression analysis controlling for various factors, such as high school, student, and social inputs, as well as other factors such as charter status and county location, provides evidence that Hispanic academic performance is

largely driven by high school and student inputs. There are significant negative relationships between the percent of English language learners at a high school and Hispanic API scores. More specifically, for a 10% increase in English language learners, Hispanic API scores are expected to decline by .02%. In addition, average class sizes in core academic courses and percent of full credentialed teachers are also found to have significant positive and negative effects on Hispanic academic performance, respectively. Results indicate that a 20% increase in class size is predicted to increase Hispanic API scores by .33%, while a 10% increase in full credentialed teachers is expected to result in a .40% decrease in Hispanic API scores. Parents, students, teachers, and the educational community at large in California must address these connections in order to better develop and prepare the Hispanic youth for future endeavors. Granted, focusing on improving these factors may prove to be difficult as in many instances administrators and educators do not have authority and control in such areas. Nevertheless, policy-makers and decision-makers must be attentive to such relationships when implementing educational strategies and reform policies.

## II. Introduction

In California, all high school students undergo identical standardized testing in order to determine academic achievement for the school overall, as well as disaggregate subgroups, such as ethnic groups, socioeconomically disadvantaged students, and students with disabilities. The performance measure utilized is the Academic Performance Index (API). The API is the California Department of Education's primary indicator in determining the various factors that have significant relationships with academic performance and academic progress in the public school system. The policy, enacted under the Public Schools Accountability Act of 1999, established a manner in which to hold schools accountable in educating students. Administrators determine API scores based on the results of standardized tests, such as the California High School Exit Exam (CAHSEE), and score each school on a scale ranging between 200-1000, with a state-

wide performance target of 800. As an improvement measure, the strategy seeks to increase academic performance of students by incentivizing schools with monetary rewards and distinguished public awards in instances that schools meet the academic growth goals. Conversely, API scores also establish which schools must undergo interventions, such as funding cuts or sanctions (California Department of Education [CDE], 2009). Furthermore, State Superintendent of Public Instruction Jack O'Connell views the API as a manner for California "to develop, implement, and sustain a specific, ambitious plan that holds the State of California accountable for creating the conditions necessary for closing the achievement gap" (CDE, 2009). Thus, the state and public schools are aware of the achievement gap and are under significant pressure to improve the institution's overall API scores, as well as API scores for various subgroups. The following research seeks to answer a research question pertaining to a specific group of high school students. More specifically, the research investigates various factors that may be correlated (and statistically significant) with changes in the API scores of Hispanic high school students in California. Three categories of explanatory variables are identified: high school, student, and social inputs and are more thoroughly described in Section IV during the discussion regarding the theoretical model.

Over the past several decades, California experienced a surge in the number of Hispanic immigrants, which now represent a significant percentage of the state's population. Consequently, Hispanics also represent the largest ethnic group in the state's public school system. During the 1998-1999 academic year, 40.5 percent of the California public school student population was Hispanic (Cheng, 2001). Intuitively, in the instance that the Hispanic population continues to increase, this ethnic group will represent the majority of California's workforce. However, Hispanics are typically found to be socioeconomically disadvantaged and in turn experience low academic performance scores. In addition to low socioeconomic status, more than one-third of Hispanic students have a parent that has not received a high school diploma (Noguchi, 2008). In the 2006 base API year, Hispanic high school students scored approximately 150 points lower than white students and nearly 200 points lower than Asian students, with African American students as the only ethnic group attaining lower scores (CDE, 2008). Granted, such social inputs are difficult for schools to address as schools do not possess the authority to intervene in the home environments of students. However a recent article in the San Diego Union Tribune, suggests that the low academic achievement of Hispanics is a result of school inputs and is a "signal that they [do not] have the same opportunity to learn because they are disproportionately herded into decrepit schools" (Moran, 2007). Such academic shortcomings of the education system, regardless of the cause, may affect future employment and earnings for Hispanics (Cheng, 2001). Thus,

one important consequence of the growing education gap, particularly for Hispanics, is a workforce skill gap that limits individuals' success and creates a significant percentage of the state's labor force that is unable to meet the future needs of California's employers. An inability to address the various issues contributing to the low academic performance of Hispanic students may also result in a lower quality of life for this ethnic group, which places them at a competitive disadvantage in comparison to other Californians. Therefore, for the purpose of this research it is imperative to assess the effects of various factors on Hispanic student achievement as measured by the API in order to identify both beneficial and problematic aspects of Hispanic students' personal and educational life as to determine possible improvement areas and effective strategies.

Figure 1 further illustrates the projected ethnic composition changes in California for between 2005 and 2025. Hispanics are forecasted to become the largest ethnic group in California, as well as the majority of the population by 2050 (PPIC, 2008).

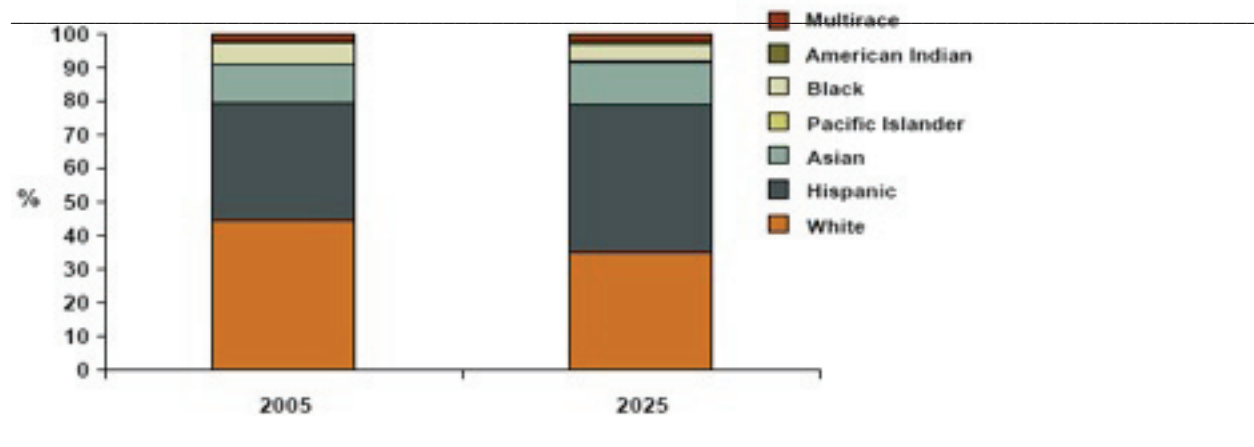
The following sections explore whether significant relationships exist between various school, student, and social inputs, and California public high school Hispanic API scores (holding all other variables constant). Section III provides a literature review which summarizes previous academic studies related to factors affecting student achievement. Section IV presents an overview of the regression model used in the analysis, including the general causal variables, descriptions for the dependent and explanatory variables, and the expected directions of each variable's effects. Section V offers details on the data sources used in the regression model, as well as descriptive statistics and the relationships between the variables. Section VI provides the regression results, which includes results from various functional forms; and possible errors in the analysis and the manners in which the errors are remedied. In the final section, section VII, the results are analyzed, giving specific attention to significant coefficients and elasticities. Also addressed are the implications of the findings on education policies and intervention strategies by school administrators in order to improve student achievement for Hispanic students, as well as suggestions for future research.

### III. Literature Review

The results of various studies indicate that certain factors relate consistently to changes in API scores. However, some factors may be out of the educational system, as well as the individual school's control (i.e. – student and social factors). Therefore, it is important to examine the type of relationship between the variables in order to assess whether public policy can in fact assist in alleviating the issues surrounding API scores. The following review summarizes research that examines various factors' effects on schools' API scores, specifically describing the purpose of the research, methods,

Figure 1

Comparison of California's Ethnic Composition: 2005 and Projected Totals for 2025



and results. The majority of the previous research assesses three general types of factors affecting student achievement: those within school control, those within school district control, and those beyond educational administrators' control. Recognizing these three different categories allows for a specific understanding of how various inputs are related to educational performance measures, such as API scores. Additionally, a table is included in Appendix A in order to further clarify researcher venues, as well as the methods and findings of the following studies.

#### Variables within School Control

##### Teacher Credentials and School Calendars' Effects on API scores

By categorizing the independent variables in three manners (i.e. – (1) socioeconomic background, (2) teachers' credentials, experience, and education; and (3) school characteristics), researchers may more specifically illustrate the different effects between factors that are within and beyond the educational system's control. Powers (2003) examined API scores for elementary school students in the two largest school districts in California, Los Angeles and San Diego Unified in this manner. The representative sample populations include 96.5% of elementary schools in Los Angeles and 95% in San Diego (Powers, 2003). The analysis utilized base data for API scores from the 1998-1999 school-year and compared it to the 2000-2001 school-year within and across each district. For the explanatory variables, the researcher created individual models according to different themes. The results of the regression analysis within each district suggest that student socioeconomic background variables, such as percentage of students qualifying for reduced-price or free lunches and percentage of English-learners, explain approximately 75% of the variability in API scores, are negatively correlated, and

more importantly statistically significant (Powers, 2003). More specifically, during 1999 in Los Angeles Unified, a one percent increase in students participating in reduced price or free lunches and in the percentage of English language learners results in a decrease of 3.15 percent and .043 percent in base API scores respectively. In 2001, the results are similar with a one percent increase in such factors resulting in decrease of 2.52 percent and .30 percent in base API scores respectively. Furthermore, during 1999, in San Diego Unified a one percent increase in students participating in reduced price or free lunches and in the percentage of English language learners results in 2.52 and .30 percent corresponding decrease in base API scores. In 2001, the magnitude of the effects remains the same for the socioeconomically disadvantage measure, while a one percent increase in English language learners results in a higher percent reduction in API scores (1.06 percent).

Moreover in Los Angeles, results indicate a negative correlation between the percentage of teachers possessing emergency credentials and API scores, while in San Diego there exists a positive correlation. The regression results in 2001 for Los Angeles indicate that a one percent increase in teachers possessing full credentials results in a 1.63 percent decrease in base API scores. Conversely for 2001, in San Diego a one percent increase in teachers possessing emergency credentials results in a 7.40 percent increase in base API scores. Such a relationship is contrary to conventional thought and brings into question the importance of teachers' credentials in influencing academic performance. In regards to school characteristics, traditional schools perform better than schools with variations on year-round calendars in both districts. It is important to note that when adding the school characteristics model to the San Diego schools' regression the correlation between teachers' credentials and API scores becomes positive, likely due to the skewed number of schools

in San Diego operating on such a calendar as there are only two. Powers' study provides evidence that factors within a school's control, such as teacher and school characteristics, may affect API scores. However, the researchers also emphasize that when considering education policy, it is dually important to consider the strong correlations between student characteristics and API scores as this suggests some factors may be beyond the schools' control.

### **Charter versus Non-Charter Schools**

Additional studies examining school characteristics assess such topics as whether the type of school (i.e. - charter or non-charter school) significantly affects API scores. Slovacek, Kunnan, & Kim (2002) examine the differences in how charter and non-charter schools in California are serving low- socioeconomic status students based on variances in API scores. Researchers analyzed API data for all California schools, as well as approximately 100 other variables, as provided by the California Department of Education for the 1999, 2000, and 2001 school years. In addition, data from California Network of Educational Charter's database was included as it provides more specific information on charter schools. Researchers focused on high-poverty schools where 50% or more of students were eligible to receive free or reduced cost lunches. The initial mean comparison between high-poverty charter and non-charter schools from 1999 to 2001 illustrates that the percent growth in API scores for charter schools (67 points) is higher than non-charter schools (64.2 points). Furthermore, when conducting a similar comparison, but solely including those schools that serve 75% of more high-poverty students, charter schools still show a larger growth in API scores (74.3 points as opposed to 68.2 points) (Slovacek et al, 2002).

Researchers also assessed the following factors when considering differences between charter and non-charter schools: percentages of English language learners, students participating in lunch programs, and teachers with full credentials. These variables create a strong predictive model for explaining 2001 API scores with an  $R = .847$  and a variance of 72%. Researchers also assessed the following factors specifically for charter schools and once again found a strong predictive model with an  $R = .81$  and a variance of 64% (Slovacek et al, 2002). The regressions indicate (for both predicting API scores for all California schools and also for solely California charter schools) that as the percentages of English language learners increases, a school's API score decreases. More specifically, a ten percent increase in the percentage of English language learners results in a 5.6 percent decrease in base API scores. The same relationship is true with the percentage of students participating in lunch programs as results indicate that a ten percent increase in the percentage of students participating in reduced price or free lunch programs is associated with a 26.5 percent decrease in base API scores. Conversely, a ten percent increase in the percentage of teachers with full credentials results in a 10.6

percent increase in base API scores suggesting that teachers with full credentials have a positive relationship with a school's API score. Future studies should continue to assess charter schools by including variables more specific to such schools, such as funding sources and educational curriculum.

### **Foreign Language Curriculum and the Effects on API Scores**

Additional exploration concerning school characteristics has led to the examination of foreign language education and such programs' effects on academic performance. Sung, Padilla, & Silva (2006) examined foreign language programs at 200 public high schools in California. A random sample of schools from 161 school districts received a survey questionnaire containing questions on various aspects of the foreign language offerings and programs, as well as background information on foreign language teachers. The regression concerning enrollment in foreign language classes and API scores, while controlling for the number of students in free or reduced lunch programs and the percentage of English- learners, yielded a positive and statistically significant correlation ( $r = .43$ ,  $p < .0001$ ) (Sung, et al, 2006). Thus, as the percent of students enrolled in foreign language classes increases, API scores increase.

Researchers suggest that as the role of foreign language education may serve as a manner in which to minimize the achievement gap between high-performing and low-performing schools, consideration for the socioeconomic status of individual students in order to determine if high enrollment in foreign language education is feasible as well as necessary. Schools with high populations of low- income students will likely not have high participation in foreign language programs as low performing schools do not have a large percentage of graduates attending colleges, which require the completion of foreign language courses. Therefore, one may conclude that variables beyond school control indirectly govern those within school control (i.e. - socioeconomic status effects on whether a student finds programs offered necessary). However, in certain instances individual public schools may not influence the decision-making process on such issues such as school district size and personnel. In these cases, the responsibility of making the respected changes in order to improve student achievement may fall on school district administrators.

### **Variables within School District Control The Effects of School District Size on API Scores**

Aside from decision-making at the individual school level, decisions made by school districts may also affect schools' API scores. For example, in California one can find various school district sizes, some of which are responsible for educating over 40,000 students and are possibly forced to spread educational resources thin (Driscoll, et. al, 2003). A study using 1999 elementary, middle-school, and high-school level API data

provided by the California Department of Education (Driscoll, et. al, 2003) examines the relationship between district size, measured by the number of students enrolled in a district, and student performance. Upon running separate regressions for each school level, researchers concluded that district size is negatively correlated and statistically significant at a 1% error level for elementary and middle-schools (Discoll, et. al, 2003). More specifically results indicate that for a one percent increase in district size, there is a 5.27 percent decrease in base API scores for elementary schools, a 4.00 percent decrease in base API scores for middles schools, and a 1.42 percent decrease in base API scores for high schools. Therefore, such results lend support to the argument that school district size plays an important role in educational program quality.

### **Correlations between School Personnel and API Scores**

Researchers also suggest that personnel services at public schools may affect student achievement, particularly the number of school counselors and school psychologists. School counselors play a role in addressing students' academic development and career exploration, while school psychologists intervene and assess students in the instance that certain events occur (Goodman & Young, 2006). A recent study conducted by Goodman and Young (2006) assesses a randomly chosen sample of 150 out of 352 public school districts from a Pacific coast state in order to determine if the number of school counselors and the number of school psychologists in a school district affected the district's composite API scores, of which the authors wrongly do not specify a year. While testing the main independent variable and controlling for socioeconomic status, researchers developed a model that explained 28% of the variance in student achievement ( $a = .05$ ). In addition, the results indicate that a one percent increase in the enrollment of socioeconomically disadvantaged students is associated with a 1.01 percent decrease in base API scores. Furthermore, the regression yielded results that indicate that the number of school psychologists has a significant positive effect on school districts' API scores ( $p < .05$ ), while the number of school counselors is positive, yet insignificant ( $p > .05$ ). More specifically, the regression results illustrate that a one unit increase in the number of school psychologists employed in the school district results in a statistically significant .69 percent increase in base API scores, while a one unit increase in the number of school counselors employed in the school district only results in a .18 percent increase in base API scores, but this effect is found to be insignificant (Goodman & Young, 2006). Goodman and Young's study suggests that administrators may choose to shift the focus from individual students to professionals within school settings and evaluate the various programs and interactions such individuals have with students. Additionally, future studies may seek to examine individual schools, as well as types of schools (i.e. – elementary, high school), as the unit of analysis in order

to obtain a more specific assessment of the relationship between individual school factors, school professionals, and API scores. Nevertheless researchers and administrators are limited in their sphere of influence and decision-making capabilities in regards to improving student performance as factors outside these individuals' control are also recognized as having significant effects on student achievement.

### **Variables beyond Educational Administrators' Control**

#### **The Effects of Race and Ethnicity**

Although researchers identify various factors associated with school characteristics, factors, such as race and socioeconomic status, exist that cannot be altered by policy-makers and educational administrators. For example, Baker et al. (2000) examine race and ethnicity in aggregate and disaggregate classifications to determine if there is an effect on math and reading performance measures. The study randomly sampled 14,596 eighth grade students from stratified samples used in the National Educational Longitudinal Study of 1988 (NELS '88).

One issue that arises in using the NELS '88 data-set is that the sample population of racial and ethnic aggregate and disaggregate groups (i.e. – White, Black, Asian, Hispanic, and Native American) are not randomly and appropriately distributed across schools, leading to difficulties in assessing whether affects on academic performance are a result of race/ethnicity, groups effects, etc.

Researchers used two sets of regression analysis to compare academic achievement differences. The first analysis compared such differences among the five aggregate groups in math and reading while controlling for socioeconomic status and language proficiency. Researchers found the performance differences to be statistically significant at the  $p < .01$  level, with Asian and White students outperforming the other groups, and observed the socioeconomic affects as strong and positive. In regards to math performance, the regression results indicate significant performance percentage decreases in scores for blacks (6.78), Native Americans (7.10), and Hispanics (3.85). Similar effects resulted in reading performance with a 3.86 percent performance decrease for black students, a 5.13 percent performance decrease for Native Americans, and a 2.01 percent performance decrease for Hispanics.

The second analysis specifically examined disaggregated subgroups in the Hispanic and Asian classifications to determine if performance differences are significant amongst subgroups and to what degree. For Hispanics no significant differences exist in regards to reading performance; however, for math, Cuban students outperform Mexican and Puerto Rican students ( $p < .01$ ). In the instance that a student is Puerto Rican, it is likely that their math performance results are 3.12 percent lower than other Hispanics. While if a student is Cuban, their math performance scores are 8.31 percent above their Hispanic counterparts. Conversely, for Asians, a significant

find indicates that Chinese students perform high in math, but perform lower than Filipino students in reading (Baker, 2000). The regression results indicate if a student is Filipino, their reading performance increases by 3.27 percent. Additionally, researchers indicate statistical evidence suggesting that language proficiency plays a role in performance differences among subgroups, while socioeconomic status does not.

This study suggests that schools may need to more specifically target ethnic groups and not aggregate race classifications. However, future studies must measure additional variables falling under student characteristics as 90% of the variance in academic achievement remained unexplained by this study (Baker, 2000). Additional variables may include generational status, gender, and reason for migration to the United States. Nevertheless, the evidence provides a rationale for tailoring academic interventions according to race and ethnicity.

#### **Various Personal Factors and API Scores**

Goe (2002) examines various variables affecting student API scores by using data from 6,387 California schools for the 1999-2000 school-year from the California Department of Education. A regression analysis indicated that multiple factors negatively correlate with API scores, such as percentage of students qualifying for free or reduced price lunches, the percentage of Hispanic students, the percentage of parents without a high school diploma, and the percentage of emergency permit teachers (Goe, 2002). More specifically, a one percent increase in students qualifying for free or reduced price lunches is associated with a 1.47 percent decrease in base API scores. Similar decreases in API scores result from a one percent increase in the percent of Hispanic students, which is associated with a .91 percent decrease in base API scores, as well as a one percent increase in the percentage of parents without a high school diploma, which results in a 1.18 percent decrease in base API scores. Lastly researchers note that a one percent increase in the percent of emergency permit teachers is associated with a .62 percent decrease in base API scores. Conversely, a positive correlation exists between the percentage of parents that attended graduate school and API scores, where a one percent increase in the percentage of parents that attended graduate school is associated with a 2.01 percent increase in base API scores. The regression also yields an adjusted R-squared suggesting that the included and statistically significant variables (mentioned above) explain a majority of the variation in test scores, providing support that many factors associated with student performance (i.e. – socioeconomic and parent demographic factors) are beyond schools' control (Goe, 2002).

#### **Key Findings from the Literature**

Evidence from the studies suggest that the majority of factors affecting students' academic achievement as reported by API scores are beyond the control of schools' and policy

makers as the factors deal with personal student characteristics. However, one must note that the studies include causal-comparative variables as the research examines changes in API scores based past "causes" (i.e. – parents' education) and thus are more prone to error if researchers have not included all possible explanatory variables. This literature review mainly focuses on studies examining student achievement within California schools. However as more research focuses on issues within the public school system and academic achievement, researchers may seek to focus on such factors as policy and curriculum differences between states in order to determine if the state's system makes a difference in educational attainment. In the instance that studies in such areas yield significant results, perhaps alterations to state's educational policies are possible and nationwide changes are necessary.

The results of the previous studies indicate that student and social factors consistently influence student achievement. For example, socioeconomically disadvantaged students, students with less educated parents, as well as minorities and English-language learners demonstrate lower student achievement. In addition, previous studies indicate school input influences, particularly in regards to the percent of fully credentialed teachers. As the following regression model focuses on different factors' effects on individual California high schools' Hispanic API scores, it is imperative to include multiple variables from the various inputs addressed in previous research in order to develop a robust model. In addition, the regression model used will not simply mirror the previous studies, but rather expand on preceding models by controlling for such inputs, such as the California county in which the high school is located, the percent of students participating in various education programs (i.e. – Gifted and Talented education programs, migrant education programs), whether the high school is a charter school, and if the institution operates on a year-round academic calendar.

#### **IV. Regression Model**

This section provides a detailed explanation of the regression model used as the basis for this analysis and includes descriptions of the dependent variable, the broad explanatory categories, the specific explanatory variables in each category, and the anticipated relationships between the explanatory variables and the dependent variable. In addition, rationales are provided for the selection of each variable.

The regression analysis is formatted after the classical regression model, also known as Original Least Squares (OLS). OLS regression is the most common regression estimation technique as the method is easy to use and minimizes residuals, which are the differences between the observed variable coefficient and the estimated regression line. The minimization of the residuals assists with the theoretical basis for the regression equation as researchers prefer for the estimated regression equation to be as near as possible

to the observed data (Studenmund, 2006). Furthermore, the research uses the OLS regression technique in an attempt to explain the variation in the dependent variable, 2007 California high school Hispanic API scores, for the chosen sample of California public high schools as correlated to the explanatory variables.

### Dependent Variable

The dependent variable in this regression model is 2007 California high school Hispanic Academic Performance Index (API) scores as it is the main measure of academic performance in the state. Education in California is major concern for education administrators. However, the motivation behind choosing a dependent variable that specifically centers on Hispanics is that the education of this ethnic group is a growing concern in California, particularly for those Hispanic students who are English language learners and are of low socioeconomic status, as Hispanics are becoming a majority of the state's student population. In the instance that the regression research can identify significant relationships between variables that "explain" decreases (and increases) in Hispanic high school students' API scores, perhaps improvements may be made in the public school system to better educate this group of students in order to increase the high school API scores and more importantly produce a better educated workforce.

### Sample

The sample frame is also an important consideration in the research design, as the number of observations included in the sample total affects the degrees of freedom and in turn the strength of the analysis. The sample population for this analysis is 995 California public high schools that reported Hispanic API scores for the 2007 academic school year. Although the total number of public high schools in California exceeds two thousand schools, not all institutions may have significant Hispanic student populations and therefore are not required to report this statistic. The current requirements for reporting subgroup API scores is a numerically significant population of either 100+ students enrolled on the first day of testing or for smaller schools, 50+ students enrolled on the first day of testing who make up a least 15 percent of the total student population. These reporting guidelines may demonstrate limitations in the assessment of the effects of various factors on Hispanic API scores in California public high schools as not all institutions may have a significant Hispanic student population. It is also important to note that the sample population chosen excludes those California public high schools that follow the Alternative Schools Accountability Model (ASAM), those that are formatted for special education, as well as high schools that are a combination of ASAM and special education.

### Theoretical Model

The theoretical model for the regression analysis focuses on factors concerned with various social characteristics, student characteristics, and school characteristics that may affect API scores. More specifically, the research question the regression analysis is seeking to answer is: What factors are correlated (and statistically significant) with changes in the API scores of Hispanic high school students in California? The dependent variable chosen is 2007 California high school Hispanic API scores, which is related to the research question in that California high school Hispanic API scores is the specific measurement the study is examining in order to determine factors affecting academic achievement for Hispanic high school students in California. Thus, the research seeks to determine if a correlation exists between increases and decreases in California Hispanic high school API scores and various factors. The potential link between the identified factors and California Hispanic public high school API scores is expressed in the following general form:

- Hispanic High School Student Achievement = f (high school inputs, student inputs, social inputs, other control variables)

where (expected effects indicated in parentheses):

- Hispanic High School Student Achievement = f [2007 Base Hispanic API score for California public high schools)
- High School Inputs = f [enrollment (?), % of teachers with full credentials (+), average class size for core academic courses (-), % students excused from testing by parents (-), % students tested (+)]
- Student Inputs = f [% African American (-), % American Indian (-), % Asian (+), % Filipino (-), % Hispanic (+), % Pacific Islander (-), % English-Language Learners (-), % Students with Disabilities (-), % Socioeconomically Disadvantaged Students (-), % Students in Gifted and Talented Education Programs (+), % Students in Migrant Education Programs (+), % Reclassified Fluent-English-Proficient Students (+)]
- Social Inputs = f [% Students on free/reduced price lunch programs (-), Parent Education: % Not High School Graduate (-), Parent Education: % High School Graduate (-), Parent Education: % Some College (+), Parent Education: % College Graduate (+), Parent Education: % Graduate School (+), Average Parent Education (+)]
- Other Control Variables = f [year round status (?), charter status (?), school size status (?), county status (?)]

### Rationale for Anticipated Effects

Hypotheses regarding the specific contributing factors within the broad general causes (that may have an effect on Hispanic high school student achievement (i.e. – A = high school inputs, B = student inputs, C = social inputs) were developed before conducting the regression research. The expected direction of these effects is indicated in parentheses in the above functional equation, where a “+” sign signifies a positive effect, a “-” sign signifies a negative effect, and a “?” sign signifies that the effect of the explanatory variable on the dependent variable is unknown. Conventional wisdom suggests that variables associated with lower socioeconomic status, which generally has a negative relationship with academic achievement, are probably also negatively associated to academic achievement and in turn base API scores. The following provides a detailed description of the variables and justification for their inclusion.

### High School Inputs

Specific factors included under the broad category of high school inputs describe particular school characteristics. Average class size for core academic courses is assumed to have a negative effect as class sizes increase it becomes more difficult for students to effectively learn as there are more distractions and students may be less attentive and in turn have low student achievement. The variable concerning the percentage of teachers with full credentials is hypothesized to have a positive relationship with Hispanic high school students’ API scores as the more teachers that have gone through a full- credential program; the more likely schools will have higher API scores due to more satisfactory and challenging curriculums. The original data set also includes a variable for the percentage of teachers with emergency credentials. However, including this variable in addition to the variable for teachers with full credentials may skew the results as the variables may cause multicollinearity. The percentage of students excused from API testing by parents is speculated to have a negative effect on Hispanic API scores as the less students present during testing, the lower the base to calculate the average API score across all Hispanic students at the specific high school. Conversely, the percentage of students tested is assumed to have a positive effect on Hispanic high school students’ API scores as the more students that participate in testing the more individual scores are able to be included in the total number of API scores possibly increasing the overall average. The final variable included in the high school input category is enrollment as the high schools in the sample population vary in size. The potential relationship of enrollment and Hispanic student achievement is unknown.

### Student Inputs

The broad category of student inputs describes student population demographics and groups. For example, the

percentage of ESL students is chosen as a specific variable under student inputs and is believed to have a negative correlation with the dependent variable as students that are not proficient in the English language will likely lower API scores due to possible decreased learning opportunities and in turn increased difficulties in understanding standardized tests. A similar hypothesis is used in predicting the effects of the percentage of students with disabilities variable. The percentage of white and Asian students at the school is also examined as it is hypothesized that Hispanics attending high schools with more white and Asian students will perform better due to desegregated learning and a more challenging environment. Furthermore, it is assumed that Hispanics attending high school with a high percentage of fellow Hispanic students will perform better on tests that affect API scores due to a more welcoming and comfortable learning environment in which Hispanic students are not outcasts. Thus, a positive correlation is predicted for the percentage of Hispanic student variable. Conversely, it is hypothesized that percentages of African American, American Indian, Filipino, and Pacific Islander students will have negative effects on Hispanic API scores as these ethnic groups are historically low academic performers and may adversely affect Hispanic students’ learning environments and lack positive educational influence. Although the last variable may be included under social factors, the percentage of socioeconomically disadvantaged students will be assessed under student characteristics as the data set includes the variable in relation to the school’s demographics. It is believed that this variable will have a negative correlation with the dependent variable as it is hypothesized that as the percentage of socioeconomically disadvantaged students decreases Hispanic high school students’ API scores will improve. Such a negative correlation may be due to the more income and occupational prestige a student’s family has, the more likely it is that parents’ may provide students with the necessary learning tools and environment outside of school.

### Social Inputs

The final broad explanatory category includes social factors. A social factor provided by the API data set is average parent education level. This variable is believed to be crucial to Hispanic high school students’ API scores as this student population may consist of a high number of first generation students. For parents that have graduated from high school and/or continued on with higher education, it is likely that these individuals encourage their children to perform well in school and believe education is important. Thus, it is hypothesized that as parents’ education level increases as will the high school’s Hispanic API scores. More detailed variables for parent education (i.e. – percentage of non-high school graduates, percentage high school graduates, percentage some college, percentage college graduates, and percentage graduate school) are included and measured as well. The same



hypothesis holds true for these variables in that as parents achieve an education level of “some college” and above, a positive relationship is assumed with Hispanic high school students’ API scores. The final social input included is the percentage of students participating in free or reduced price lunch programs. Similar to the socioeconomic disadvantaged student input variable, this variable is likely to have a negative effect on Hispanic high school API scores in that students who participate in such programs are likely to be of lower socioeconomic status, which in turn may suggest a less conducive home environment for academic achievement.

### Other Control Variables

Other control variables (i.e. – year round status, charter status, school size status, and county status) are also included in the analysis and hypothesized to have unknown effects on the dependent variable. For example, the year-round dummy variable (coded 1 in the instance that the high school is year-round and coded 0 otherwise) may be hypothesized that if the high school is year-round, students will experience less of a decrease in achievement scores as schools eliminate the three-month summer break and provide students with a more consistent academic schedule. Conversely, non-year round high schools may also have a positive effect on API scores as it may be beneficial to Hispanic student achievement for students to have a longer vacation period during the academic year. Such a break period may allow students to return to an educational setting more rejuvenated leading to higher attentiveness and in turn higher test scores.

### V. Data

In order to effectively interpret the results of the regression analysis, it is necessary to thoroughly explain the explanatory variables used in the model. Thus, this section provides general clarifications of the included variables. Additional information may also be found in the series of charts illustrating the data sources for the variables, as well as descriptive statistics and the correlation coefficients between each of the variables.

### Explanatory Variables

Explanatory variables are divided into three separate input categories: school, student, and social inputs. The category concerning school factors includes variables such as total enrollment, percentage of teachers possessing full credentials, average class size for core academic courses, percentages of students excused from testing by parents, and percentage of students tested. For student and social factors, the included explanatory variables are used as control variables in the majority of previous research and reflect major factors considered to have significant relationships to student achievement. Variables such ethnic background, stated as percentage of the student population included in each ethnic group, the percentage of students participating

in various education programs offered at the high school, the percentage of socioeconomically disadvantaged students, percentage of English-learners, and percentage of students with disabilities are included in the student input category. For social characteristics, the factors controlled for are percentage of students participating in free and reduced price lunch programs and an aggregate measure of parent education expressed as a value on a 1 to 5 scale. Disaggregate measures of parent education groups (i.e. – not high school graduate, high school graduate, etc.) are also included in order to further assess the specific effects of such groups on Hispanic high school student achievement. Additional controls, such as dummy counties variables, are included in order to draw comparisons amongst counties. It is important to note in regards to the county dummy variables, only 57 counties are shown in the statistics as the Los Angeles County Dummy is excluded from the regression results in order to decrease the probability of erroneous results; thus, establishing Los Angeles County as the base of comparison for all other county dummy variables.

### Data Sources and Descriptive Statistics

The following tables contain additional details on the explanatory variables, specifically descriptions and sources (Table 1), descriptive statistics, including the mean, standard deviation, maximum and minimum values (Table 2), and bivariate correlation coefficients between the explanatory variables, which illustrate the strength and direction of the linear relationships (Table 3: See Appendix B).

### VI. Regression Analysis

This section presents the results of the regression analysis performed on the data set. A comparison of the various functional forms is outlined in Table 4, which is followed by a justification for the functional form selected, the processes used in order to find errors in the regression analysis and the manners in which such errors are remedied.

### Selecting a Functional Form

Table 4 illustrates how regression findings differ based on the functional form used. The log-log form is chosen as the functional form as the log-log regression results yielded 28 variables significant at the 90% confidence level. Conversely, the linear and log-linear regressions have only 22 variables and 21 variables significant at the 90% confidence level respectively. The expected directions for the following variables were found significant and in the expected direction in all three functional forms: percentage of students participating in free/reduced priced lunch programs (-), percent Pacific Islander (-), percent English learners (-), percent in Gifted and Talented Education programs (+), percent reclassified fluent English proficient students (+). However, surprisingly the results for all functional forms indicate that the percent of socioeconomically disadvantaged students

Table 1  
Variable Labels, Descriptions, and Data Sources

Variable Label	Description	Source
<b>Dependent Variable</b>		
HI_API07	2007 California high school's Hispanic API Scores (Base Score)	California Department of Education (CDE), 2008 Growth API Data File; <a href="http://www.cde.ca.gov/ta/ac/ap/api-datafiles.asp">http://www.cde.ca.gov/ta/ac/ap/api-datafiles.asp</a>
<b>Independent Variables: High School Inputs</b>		
Enrollment	Number of students enrolled on first day of testing	CDE, 2008 Growth API Data File
Full Credential	Percent of teachers with full credentials	CDE, 2008 Growth API Data File
Class Size	Average Class Size for Core Academic Courses	CDE, 2008 Growth API Data File
Excused	Percent of students excused from testing by parents	CDE, 2008 Growth API Data File
Tested	Percent of students tested	CDE, 2008 Growth API Data File
<b>Independent Variables: Student Inputs</b>		
African American	Percent of African American students	CDE, 2008 Growth API Data File
American Indian	Percent of American Indian students	CDE, 2008 Growth API Data File
Asian	Percent of Asian American students	CDE, 2008 Growth API Data File
Filipino	Percent of Filipino students	CDE, 2008 Growth API Data File
Hispanic	Percent of Hispanic students	CDE, 2008 Growth API Data File
Pacific Islander	Percent of Pacific Islander students	CDE, 2008 Growth API Data File
English-language learners	Percent of students English- language learners	CDE, 2008 Growth API Data File
Disabilities	Percent of students with disabilities	CDE, 2008 Growth API Data File
Socioeconomically Disadvantage	Percent of socioeconomically disadvantaged students	CDE, 2008 Growth API Data File
GATE	Percent of students in Gifted and Talented Education programs	CDE, 2008 Growth API Data File
Migrant Education	Percent of students in Migrant Education programs	CDE, 2008 Growth API Data File
Reclassified	Percent of Reclassified Fluent English Proficient students	CDE, 2008 Growth API Data File

Table 1  
Variable Labels, Descriptions, and Data Sources

Variable Label	Description	Source
<b>Independent Variables: Social Inputs</b>		
Free/Reduced Lunches	Percent of students participating in free/reduced priced lunches	CDE, 2008 Growth API Data File
Parent Education Level	Parent's average education level (5 point scale in which 1=no high school education and 5=graduate school)	CDE, 2008 Growth API Data File
Not High School Grad	Parent Education Level: Percent Not High School Graduate	CDE, 2008 Growth API Data File
High School Grad	Parent Education Level: Percent High School Graduate	CDE, 2008 Growth API Data File
Some College	Parent Education Level: Percent Some College	CDE, 2008 Growth API Data File
College Grad	Parent Education Level: Percent College Graduates	CDE, 2008 Growth API Data File
Grad School	Parent Education Level: Percent Graduate School	CDE, 2008 Growth API Data File
<b>Independent Variables: Additional Controls</b>		
Dummy: Year Round?	Dummy variable for status as a year round high school (coded 1 = year round; 0 = otherwise)	CDE, 2008 Growth API Data File
Dummy: Charter?	Dummy variable for status as a charter school (coded 1 = charter; 0 = otherwise)	CDE, 2008 Growth API Data File
Dummy: Small School?	Dummy variable for status as a small school for 2007 API (coded 1 = small school; 0 otherwise)	CDE, 2008 Growth API Data File
County Names	Dummy variables for county status; total of 57 dummies for the represented counties in the data set. (coded 1 = if high school represents county; 0 = otherwise). Los Angeles County is the excluded dummy and is the base for comparison.	CDE, 2008 Growth API Data File

Table 2  
Descriptive Statistics

Variable Label	Mean	Standard Deviation	Minimum Value	Maximum Value
<b>Dependent Variable</b>				
HI_API07	640.12	76.991	368	930
<i>Independent Variables: High School Inputs</i>				
Enrollment	1255.82	779.99	55	3856
Full Credential	91.15	10.89	9	100
Class Size	26.40	5.03	4	40
Excused	.18	.42	0	3.95
Tested	97.97	3.01	60.99	100
<b>Independent Variables: Student Inputs</b>				
African American	18.71	15.28	3	62
American Indian	.36	.479	0	1
Asian	15.70	13.90	0	50
Filipino	2.76	4.41	0	46
Hispanic	51.04	25.46	6	99
Pacific Islander	.63	.96	0	8
English-language learners	19.27	14.814	0	100
Disabilities	9.79	11.84	0	100
Socioeconomically Disadvantage	.02	23.88	.87	99.69
GATE	10.11	9.47	0	75
Migrant Education	2.58	5.95	0	44
Reclassified	17.97	12.49	0	65
<b>Independent Variables: Social Inputs</b>				
Free/Reduced Lunches	46.91	26.01	0	100
Parent Education Level	2.60	.624	1	5
Not High School Grad	24.64	17.12	0	100
High School Grad	25.77	10.07	0	100
Some College	22.30	9.03	0	50
College Grad	17.55	10.98	0	100
Grad School	8.88	9.275	0	100
<b>Independent Variables: Additional Controls</b>				
Dummy: Year Round?	.021	.144	0	1
Dummy: Charter?	.040	.197	0	1

Table 2  
Descriptive Statistics

Variable Label	Mean	Standard Deviation	Minimum Value	Maximum Value
Dummy: Small School?	.049	.216	0	1
Alameda County	.030	.182	0	1
Butte County	.004	.063	0	1
Colusa County	.004	.063	0	1
Contra Costa County	.019	.137	0	1
El Dorado County	.001	.032	0	1
Fresno County	.040	.197	0	1
Glenn County	.003	.055	0	1
Humboldt County	.001	.032	0	1
Imperial County	.012	.109	0	1
Inyo County	.001	.032	0	1
Kern County	.024	.154	0	1
Kings County	.005	.071	0	1
Lake County	.004	.063	0	1
Los Angeles County	.240	.427	0	1
Madera County	.007	.084	0	1
Marin County	.002	.045	0	1
Mendocino County	.004	.063	0	1
Merced County	.013	.114	0	1
Modoc County	.001	.032	0	1
Mono County	.001	.032	0	1
Monterey County	.014	.118	0	1
Napa County	.005	.072	0	1
Orange County	.069	.254	0	1
Placer County	.007	.084	0	1
Riverside County	.062	.242	0	1
Sacramento County	.042	.201	0	1
San Benito County	.003	.055	0	1
San Bernardino County	.059	.236	0	1
San Diego County	.094	.293	0	1
San Francisco County	.017	.130	0	1
San Jose County	.021	.144	0	1
San Luis Obispo County	.007	.084	0	1
San Mateo County	.019	.137	0	1

Table 2  
Descriptive Statistics

---

Variable Label	Mean	Standard Deviation	Minimum Value	Maximum Value
Santa Barbara County	.011	.105	0	1
Santa Clara County	.036	.187	0	1
Santa Cruz County	.006	.077	0	1
Solano County	.012	.109	0	1
Sonoma County	.017	.130	0	1
Stanislaus County	.019	.137	0	1
Sutter County	.003	.055	0	1
Tehama County	.001	.032	0	1
Tulare County	.024	.154	0	1
Ventura County	.020	.140	0	1
Yolo County	.005	.071	0	1
Yuba County	.003	.144	0	1

---

Table 3  
Regression Results Across Functional Forms

Variable Label Ln=Variable in log form	Log – Log	Log – Lin	Linear	Fs for Log – Log
Constant	2.671 (.021)	5.213 (.000)	-68.222 (.910)	
<i>Independent Variables: High School Inputs</i>				
Enrollment (Ln)	.000 (.908)	4.972E-6 (.319)	-.006* (.055)	2.370
Full Credential (Ln)	-.040** (.046)	7.154E-7 (.998)	-.013 (.944)	1.433
Class Size (Ln)	.033** (.033)	.001 (.416)	.173 (.708)	1.892
Excused	.003 (.743)	.004 (.645)	1.277 (.784)	1.357
Tested (Ln)	.692*** (.000)	.006*** (.000)	3.581*** (.000)	1.331
<i>Independent Variables: Student Inputs</i>				
African American	-9.245E-5 (.660)	-7.999E-5 (.688)	-.026 (.829)	1.302
American Indian	-.001 (.816)	.000 (.950)	.386 (.917)	1.170
Asian	.000 (.502)	.000 (.423)	.071 (.579)	1.178
Filipino	.001 (.415)	.001 (.103)	.853* (.067)	1.610
Hispanic (Ln)	-.020** (.046)	.000 (.271)	.180 (.257)	5.523
Pacific Islander	-.010*** (.010)	-.009*** (.009)	-5.296** (.015)	1.629
English-language learners	-.002*** (.000)	-.002*** (.000)	-1.530*** (.000)	4.002
Disabilities	.000 (.353)	1.341E-5 (.975)	.138 (.599)	1.543
Socioeconomically Disadvantage (Ln)	.032*** (.000)	.004*** (.000)	2.089*** (.000)	5.528
GATE	.001*** (.005)	.001*** (.003)	.727*** (.001)	1.683
Migrant Education	.002*** (.000)	.001** (.040)	.723* (.068)	2.079
Reclassified	.002*** (.000)	.001*** (.009)	.607*** (.007)	2.610
<i>Independent Variables: Social Inputs</i>				

Table 3  
Regression Results Across Functional Forms

Variable Label Ln=Variable in log form	Log – Log	Log – Lin	Linear	Fs for Log – Log
Free/Reduced Lunches	-.001*** (.000)	.003*** (.000)	-1.581*** (.000)	6.483
Parent Education Level	.458 (.146)	.424 (.155)	241.954 (.186)	4930.660
Not High School Grad	.002 (.808)	.001 (.884)	.467 (.916)	2136.826
High School Grad	-.003 (.547)	-.003 (.518)	-2.189 (.484)	352.804
Some College	-.006 (.175)	-.006 (.177)	-3.421 (.194)	203.004
College Grad	-.010* (.072)	-.009* (.090)	-5.118 (.116)	478.081
Grad School	-.014* (.068)	-.013* (.083)	-7.315 (.111)	688.250
<i>Independent Variables: Additional Controls</i>				
Dummy: Year Round?	-.047** (.027)	-.034* (.092)	-20.315 (.103)	1.202
Dummy: Charter?	.011 (.473)	.017 (.231)	11.067 (.209)	1.075
Dummy: Small School?	-.033* (.053)	.002 (.917)	.132 (.989)	1.680
Alameda County	-.058*** (.001)	-.056*** (.001)	-36.936*** (.000)	1.385
Butte County	-.064 (.162)	-.060 (.167)	-43.072 (.103)	1.085
Colusa County	.000 (.993)	-.002 (.963)	-7.619 (.774)	1.098
Contra Costa County	-.020 (.366)	-.012 (.561)	-10.558 (.411)	1.196
El Dorado County	-.006 (.950)	-.022 (.796)	-21.675 (.680)	1.072
Fresno County	-.016 (.354)	-.009 (.571)	-3.585 (.723)	1.463
Glenn County	.045 (.403)	.046 (.361)	27.984 (.366)	1.112
Humboldt County	-.041 (.648)	-.023 (.781)	-21.652 (.674)	1.044
Imperial County	-.036 (.228)	-.023 (.426)	-12.376 (.480)	1.288



Table 3  
Regression Results Across Functional Forms

Variable Label Ln=Variable in log form	Log – Log	Log – Lin	Linear	Fs for Log – Log
Inyo County	-.119 (.184)	-.113 (.181)	-78.732 (.129)	1.042
Kern County	-.037* (.069)	-.031 (.107)	-22.807* (.054)	1.275
Kings County	-.015 (.722)	-.014 (.718)	-13.567 (.573)	1.123
Lake County	-.067 (.154)	-.059 (.182)	-43.915 (.103)	1.134
Madera County	.040 (.286)	.049 (.164)	26.386 (.222)	1.080
Marin County	-.026 (.688)	-.027 (.658)	-26.601 (.473)	1.066
Mendocino County	.003 (.953)	.008 (.858)	-.330 (.990)	1.104
Merced County	-.043* (.095)	-.030 (.224)	-18.111 (.227)	1.114
Modoc County	.053 (.555)	.050 (.558)	22.984 (.660)	1.054
Mono County	-.055 (.539)	-.073 (.385)	-55.506 (.283)	1.043
Monterey County	-.090*** (.001)	-.073*** (.005)	-45.630*** (.004)	1.225
Napa County	-.047 (.247)	-.033 (.385)	-25.368 (.282)	1.074
Orange County	.032** (.019)	.053*** (.000)	32.651*** (.000)	1.435
Placer County	-.048 (.175)	-.034 (.315)	-26.158 (.200)	1.147
Riverside County	-.027** (.046)	-.019 (.136)	-11.994 (.133)	1.330
Sacramento County	.003 (.851)	.017 (.301)	7.637 (.449)	1.625
San Benito County	-.141*** (.009)	-.112** (.029)	-68.241** (.029)	1.137
San Bernardino County	.012 (.393)	.007 (.616)	4.371 (.605)	1.479
San Diego County	-.029** (.015)	.025** (.031)	-18.815*** (.007)	1.544
San Francisco County	-.030 (.258)	.008 (.746)	-6.213 (.684)	1.567

Table 3  
Regression Results Across Functional Forms

Variable Label Ln=Variable in log form	Log – Log	Log – Lin	Linear	Fs for Log – Log
San Jose County	-.049** (.023)	-.038* (.065)	-25.357** (.043)	1.243
San Luis Obispo County	-.007 (.849)	-.002 (.942)	-7.450 (.716)	1.136
San Mateo County	-.028 (.260)	.021 (.370)	-22.057 (.131)	1.510
Santa Barbara County	-.038 (.190)	-.032 (.241)	-24.783 (.140)	1.089
Santa Clara County	-.052*** (.002)	-.044*** (.005)	-32.605*** (.001)	1.235
Santa Cruz County	-.138*** (.000)	-.121*** (.001)	-82.744*** (.000)	1.104
Solano County	-.040 (.153)	-.028 (.290)	-22.944 (.159)	1.232
Sonoma County	-.069*** (.005)	-.062*** (.009)	-45.410*** (.002)	1.183
Stanislaus County	.006 (.790)	.022 (.314)	12.169 (.357)	1.245
Sutter County	.001 (.988)	.008 (.878)	-.392 (.990)	1.058
Tehama County	-.019 (.832)	-.028 (.735)	-22.322 (.664)	1.021
Tulare County	-.015 (.491)	-.012 (.545)	-11.382 (.368)	1.266
Ventura County	.004 (.860)	.010 (.630)	5.145 (.683)	1.211
Yolo County	-.069* (.087)	-.061 (.108)	-44.521* (.058)	1.061
Yuba County	-.010 (.842)	.008 (.870)	-.922 (.976)	1.077
<b>R-Squared</b>	.544	.592	.590	
<b>Adjusted R-Squared</b>	.508	.560	.558	
<b>Observations</b>	963	963	963	

Shaded Column: Functional Form Used

Ln Variables = Logged variables (variables that do not have zero or negative values)

\*significant at the 90% confidence level, \*\*significant at the 95% confidence level,

\*\*\*significant at the 99% confidence level (All tests are two-tailed t-tests.)

Note: Los Angeles County Dummy excluded from regression results to decrease chances of erroneous results. Therefore, Los Angeles County is the base of comparison for all other county dummies.

Table 4  
VIF Comparison: Pre- and Post-Corrections for Multicollinearity

Variable Label Ln = Variable in log form	VIFs Uncorrected for Multicollinearity	VIFs Corrected for Multicollinearity
<i>Independent Variables: High School Inputs</i>		
Enrollment (Ln)	2.370	2.38
Full Credential (Ln)	1.433	1.419
Class Size (Ln)	1.892	1.841
Excused	1.357	1.345
Tested (Ln)	1.331	1.305
<i>Independent Variables: Student Inputs</i>		
African American	1.302	1.305
American Indian	1.170	1.166
Asian	1.178	1.181
Filipino	1.610	1.601
Hispanic (Ln)	5.523	4.988
Pacific Islander	1.629	1.614
English-language learners	4.002	3.888
Disabilities	1.543	1.519
Socioeconomically Disadvantage (Ln)	5.528	*Excluded*
GATE	1.683	1.662
Migrant Education	2.079	2.075
Reclassified	2.610	2.523
<i>Independent Variables: Social Inputs</i>		
Free/Reduced Lunches	6.483	4.394
Parent Education Level	4930.660	*Excluded*
Not High School Grad	2136.826	9.210
High School Grad	352.804	2.866
Some College	203.004	2.596
College Grad	478.081	3.880
Grad School	688.250	3.283

has a significant and positive relationship with Hispanic high school API scores.

Furthermore, the Adjusted R-Squared values are higher in both the linear and log-linear regressions. However, this value is not comparable across the various functional forms. Thus, the choice on which functional form to utilize is based upon the number of significant relationships and in turn the log-log functional form is chosen. The double log functional form is interpreted as a percent change in an explanatory variable (holding all other variables constant) results in a percent change in the dependent variable (Studenmund, 211).

### Errors in Regression Results

The most important phases in regression analysis are specifying a theoretically sound regression model, which includes the careful selection of explanatory variables, indicating the manners in which these variables are measured, as well as choosing the most effective functional form to illustrate the relationship between the independent variables and the dependent variable. Although careful consideration may be taken in order avoid various errors that result from steps involved in specifying a regression equation, certain errors may arise nonetheless as such errors naturally occur with the specific data set. Two common issues, particularly found in cross-sectional data models in which observations are collected from the same time frame (i.e. – 2007 academic school year), but are collected from different units (i.e. – high schools), are multicollinearity and heteroskedasticity. The following discussion addresses whether these issues are present in the regression results, and identifies the manners in which the problems may be and are remedied.

#### Multicollinearity

Multicollinearity is a violation of the Classical Assumption which states that, “no explanatory variable is a perfect linear function of any other explanatory variables” (Studenmund, 246). In the instance that there is a strong linear relationship between two or more explanatory variables, it is extremely difficult to assess the individual variable’s affects on the dependent variable. Such strong relationships may result from the specific sample chosen or theoretical errors. For example, research surrounding social science topics that typically include conditional variables such as socioeconomic status, education level, and English-language capabilities, may often experience high multicollinearity as these variables are likely interrelated. Nevertheless, it is important to consider that it is extremely rare to observe a regression equation in which none of the explanatory variables are correlated with one another. Thus, multicollinearity is expected. However, it is the degree to which multicollinearity exists in the equation that is important to note. Although multicollinearity does not adversely affect the regression coefficients by creating bias, it may lead to high standard errors and low t-scores which in turn create difficulties in achieving statistical significance.

Generally, two techniques are used in detecting

multicollinearity. The first involves the examination of the bivariate correlation coefficients which indicate the strength and direction of the relationship between two variables. In instances where correlation coefficients are high in absolute value, typically greater than 0.80, there is an indication of serious multicollinearity. As a result these values and relationships are identified (as noted in Table 3, Appendix B). Furthermore, multicollinearity may be detected by calculating the variance inflation factors (VIFs) for each explanatory variable (see Table 4). This measurement assesses the degree to which one explanatory variable may be explained by all other explanatory variables in the equation. The general rule states that VIFs greater than 5 indicate significant multicollinearity.

Results for this regression analysis, exhibit bivariate correlation coefficients greater than 0.80 and VIFs greater than 5 (Studenmund, 257-258). Such indications of multicollinearity are observed in relationships between student and social inputs, as well as relationships between two social inputs. In some instances (i.e. – the high bivariate correlation coefficient found between percentage of socioeconomically disadvantaged students included in the API and percentage of students participating in free or reduced price lunches), it may be necessary to eliminate either of the variables as the variable may be measuring the same effects. Typically students that are socioeconomically disadvantaged are participating in free or reduced price lunches and in turn the data may be repetitive. Although, this may not be true in all cases and arguments may be made on a theoretical level to include both variables in the regression equation as the variables are pertinent to the purpose of the regression analysis: to include and control for as many student, school, and social inputs in order to assess various variables’ affects on 2007 Hispanic high school API scores, a decision was made to exclude percentage of socioeconomically disadvantaged students in order to correct for multicollinearity. In this instance, it is wise to drop a redundant variable (Socioeconomically Disadvantaged), which was initially included in order to prevent omitted variable bias, but now found to represent a similar effect on the dependent variable as another explanatory variable. Similarly, high correlation coefficients exist between the aggregate and disaggregate variables for parent education. In order to observe, more detailed relationships between the various levels of parent education and a high school’s Hispanic API score, the aggregate variable measuring average parent education level on a scale is excluded. Thus, the main remedy used to reduce multicollinearity in this regression model is the exclusion of two redundant variables (Socioeconomically Disadvantaged and Parent Education Level).

#### Heteroskedasticity

Heteroskedasticity also violates one of the Classical Assumptions, which states that “observations of the error term are drawn from a distribution that has a constant variance” (Studenmund, 346). Thus, the condition of

heteroskedasticity is typically found in cross-sectional data sets where inconsistencies exist in the variance of the error term between larger and smaller observations (i.e. – large high schools and small high schools). Similar to multicollinearity, the consequences of heteroskedasticity do not lead to bias in the estimated coefficients. Nonetheless, heteroskedasticity adversely affects the regression analysis by likely leading to OLS incorrectly estimating the true coefficients, as well as the standard errors, which leads to an overestimation of the t-scores making them unreliable when hypothesis testing. For example, if a correlation coefficient's t-score is too high, it increases the likelihood that one will reject a null hypothesis when in fact it cannot be rejected. Thus, heteroskedasticity increases the chance of committing Type I errors.

A main method in testing for heteroskedasticity is the Park test, in which the residuals (or estimates of the error terms) from the regression are squared, logged, and tested in a subsequent regression. The second regression uses the log of the square residuals as the dependent variable and a Z factor, or the log of an explanatory variable (which may or may not be found in the original regression equation) that appears to vary with variance of the error term. When testing for heteroskedasticity in this analysis, the researcher identifies the variable "enrollment on the first day of testing" as the Z factor since there is suspect of significant variation in size of high schools within the data set. The interactive scatterplot of the squared residuals plotted against the Z factor (see Appendix C) illustrates that heteroskedasticity is not present. Additionally, the results of the t-test are not statistically significant indicating the possibility that heteroskedasticity is not present in the equation and that enrollment and the residuals are not related. Furthermore, the results suggest that one may not reject the null hypothesis of homoskedasticity as the absolute value of the calculated t-score (.963 or |-.963|) is less than the critical t-value (2.576 at a 1% level of confidence for a two-tailed t-test with 892 degrees of freedom). Thus, there are no corrections necessary for heteroskedasticity.

## VII. Conclusion

The final regression results are summarized in the following subsections, which include discussions of the model fit, if the results of the study reflect prior expectations, the relevance of the significant variables in relation to Hispanic API scores, as well as potential policy implications and suggestions for future research on the topic.

### Elasticities and Confidence Intervals

As demonstrated in Table 4, some of the variables included in the log-log functional form are able to be expressed in log form whereas others are not. When drawing conclusions regarding the relationships between the explanatory variables and California high school Hispanic API scores, one may conclude that either a unit or percentage change in the explanatory variable results in a unit or percentage change

in the dependent variable. However, the coefficients are converted into elasticities in order to more easily compare the variables by addressing the magnitude of the effect of each explanatory variable on the dependent variable as the percentage change in the dependent variable given a one percent change in the explanatory variable (when holding all other factors constant). Furthermore, the coefficients and elasticities are converted into confidence intervals at the 90 percent level, which illustrate the range of possible values. Table 6 includes the coefficients and elasticities, as well as the confidence intervals for all the significant variables at the 90% level. The calculations used to compute each of the values may be found following the table.

## Discussion of Findings

### Model Fit

When evaluating the findings of a regression analysis, it is important to examine the quality of the regression equation and how well the model fits the data. The coefficient of determination, or R-squared, is observed as it indicates that percentage of variation in the dependent variable that is explained by the regression equation (Studenmund, 50). The range for the R-squared values is zero to one, with values closer to zero indicating a poor model and values closer to one indicating a "better fit." The R-squared for the log-log regression used in this study is .544 (as illustrated in Table 6). This indicates that the regression model explains slightly more than the majority of the variation in Hispanic API scores, which may be considered satisfactory. However, as the R-squared value remains considerably lower than one, which indicates a perfect model fit, the regression equation likely excludes specific variables that may be relevant to differences in California high school Hispanic API scores. Thus, future analyses may take into consideration factors, such as school district size, school district personnel, curriculum, or factors associated with individual tests that are used to determine API scores. This suggests the possibility of omitted variable bias, which indicates that the interpretation of the results must bear in mind that some factors associated with variance in high school Hispanic API scores are excluded, which in turn implies that the results must be interpreted with caution noting that as a portion of the variance is random or unexplained by the model.

### Relationship between Expectations and Results of Significant Variables

The relationship between the expected directions of the explanatory variables (when holding all other variables constant) and the results from the regression analysis illustrate whether the study's findings are consistent with previous literature or may contribute additional information to the research. Many of the study's statistically significant results are consistent with previous studies concerning various factors effects on API scores, while others question

Table 5  
Elasticities and Confidence Intervals for Significant Variables

Variable Label Ln = Variable in log form	Best OLS Results Log – Log	Elasticity	Confidence Level Range (90% level)
Constant	2.671 (.021)		
<i>Independent Variables: High School Inputs</i>			
Full Credential (Ln)	-.040** (.046)	-.040	-.092 to .012
Class Size (Ln)	.033** (.033)	.033	-.006 to .072
Tested (Ln)	.692** (.000)	.692	.44 to .944
<i>Independent Variables: Student Inputs</i>			
Hispanic (Ln)	-.020** (.046)	-.020	-.049 to .009
Pacific Islander	-.010*** (.010)	-.0010	-.02 to .00
English-language learners	-.002*** (.000)	-.006	-.002 to -.002
Socioeconomically Disadvantage (Ln)	.032*** (.000)	.032	.009 to .055
GATE	.001*** (.005)	.002	.001 to .001
Migrant Education	.002*** (.000)	.0008	-.001 to .005
Reclassified	.002*** (.000)	.006	.002 to .002
<i>Independent Variables: Social Inputs</i>			
Free/Reduced Lunches	-.001*** (.000)	-.007	-.001 to -.001
College Grad	.010* (.072)	-.027	-.026 to .006
Grad School	-.014* (.068)	-.019	-.035 to .007
<i>Independent Variables: Additional Controls</i>			
Dummy: Year Round?	-.047** (.027)	-.0002	-.101 to .007
Dummy: Small School?	-.033* (.053)	-.0003	-.077 to .011
Alameda County	-.058*** (.001)	-.0003	-.104 to -.012
Kern County	-.037* (.069)	-.0001	-.089 to .015

Table 5  
Elasticities and Confidence Intervals for Significant Variables

Variable Label Ln = Variable in log form	Best OLS Results Log – Log	Elasticity	Confidence Level Range (90% level)
Merced County	-.043* (.095)	-.00007	-.110 to .024
Monterey County	-.090*** (.001)	-.0001	-.160 to -.020
Orange County	.032** (.019)	.0004	-.004 to .068
Riverside County	-.027** (.046)	-.0003	-.063 to .009
San Benito County	-.141*** (.009)	-.00007	-.281 to -.001
San Diego County	-.029** (.015)	-.0004	-.060 to .002
San Jose County	-.049** (.023)	-.0002	-.106 to .008
Santa Clara County	-.052*** (.002)	-.0003	-.096 to -.008
Santa Cruz County	-.138*** (.000)	-.0002	-.236 to -.040
Sonoma County	-.069*** (.005)	-.0002	-.133 to -.005
Yolo County	-.069* (.087)	-.0001	-.172 to .034
<b>R-Squared</b>	.544		
<b>Adjusted R-Squared</b>	.508		
<b>Observations</b>	963		

\*Significant at the 90% confidence level

\*\*Significant at the 95% confidence level

\*\*\*Significant at the 99% confidence level (All two-tailed t-tests)

Formulas Used:

Elasticity = Coefficient  $i$  \* (mean of  $X_i$ /mean of  $Y$ ): Coefficient  $i$  = Elasticity if variable is in log-form

Confidence Interval = Coefficient  $i$  +/- (Standard Error  $i$ \* Critical  $t$ ); Critical  $t$  = 2.576

the explanatory variables expected relationships. For example, the student factor, percent of English-language learners demonstrates the expected negative relationships to high school Hispanic API scores (significant at the 99% confidence level). In terms of additional social factors showing statistical significance at the 99% confidence level and the 90% confidence level respectively, percent of students participating in free or reduced price lunch programs, percent of parents who are college graduate, and percent of parents who attended graduate school have negative relationships with API scores. These results concerning parent education levels are contrary to the expectations as one would believe that a higher population of educated parents may lead to higher API scores. However, the difference in the findings may be explained as the parents who are college graduates or have attended graduate school are not parents of Hispanic students. Therefore, these students do not have the benefits of individuals with experience in higher education at home to positively influence their schooling to create a conducive learning environment. Thus, a negative relationship exists.

In regards to ethnicity, percent Pacific Islander is negatively related as expected. However, percent Hispanic is also found to be statistically significant and negatively related, which is contrary to the expectations. A positive relationship was expected as Hispanic students attending high schools with high Hispanic populations may be more likely to perform well in a more comfortable learning environment where there are a majority of students with similar backgrounds and experiences and the students do not feel excluded. On the other hand, a negative relationship may exist due to the fact that Hispanics generally have low academic scores and the higher percentage of Hispanics at a high school; the increase likelihood that poor performing Hispanic students are contributing to the API score. Furthermore, percent of students in Gifted and Talented Education Programs, migrant education programs, and percent of students reclassified as Fluent-English-Proficient students have significantly positive effects on high school Hispanic API scores.

The findings concerning two of the included school inputs, full credentialed teachers and class size in core academic courses are inconsistent with the expected relationships. For example, percent of full credentialed teachers is found to be statistically significant, but has a negative relationship to high school Hispanic API scores. Conversely, class size in core academic courses is hypothesized to have a negative relationship with the dependent variable, but the regression indicates a positive relationship. The results show a small relationship for both variables at the 95% confidence level. For percent of full credentialed teachers, the results may suggest that it is not necessarily the type of teaching credential a teacher possesses that may affect Hispanic API scores (as the results of this study and previous studies conflict), but rather other factors must be considered, such as race/ethnicity of the teacher. Moreover, the positive effect of class size in core

courses may indicate that Hispanic perform better when engaged with more students in the classroom. However, both factors must be further examined in order to better determine the reasons for the unexpected relationships.

In regards to other control variables, which were not predicted to have any specific effects, the year-round and small school dummy variables are significant and are negatively related to Hispanic API scores indicating that, when holding all other variables constant, Hispanics score lower on the API scale when attending high schools operating on a year-round academic calendar or when attending small schools, which report only 11-99 valid API scores (CDE, 2009). Furthermore when controlling for individual counties, the regression yielded significant negative relationships for the following counties: Alameda, Kern, Monterey, Riverside, San Benito, San Diego, San Jose, Santa Clara, Santa Cruz, Sonoma, and Yolo; and only one statistically significant positive relationship for Orange county. These results suggest that specific counties (and possibly school districts) must be further examined in detail in order to determine why Hispanic high school students perform better in specific counties versus others.

#### **Magnitude of the Effects of Significant Factors**

The relative impact of each explanatory variable on the dependent variable is observed in the magnitude of the elasticities. For example, an explanatory variable that has a high elasticity is more influential than an explanatory variable with a lower elasticity. Therefore when examining the results of research, it is important to focus on the explanatory variables that are more likely to greatly impact the dependent variable. Consequently, the development and implementation of policies and reforms must primarily concern such variables as these factors will yield the most benefits or assist in alleviating great costs.

In this study, factors that significantly impact test scores according to the elasticities are the following: percent full credentialed teachers, class size in core academic courses, percent Hispanic students, and percent of students tested for the API. More specifically, for a 10% increase in full credentialed teachers, Hispanic API scores are expected to decline by .40%. The impact of class size in core academic courses is similar, but positive, as a 10% increase in class size is predicted to increase Hispanic API scores by .33%. Furthermore, a 10% increase in Hispanic students is expected to decrease Hispanic API scores by .20%. Lastly, a 10% increase in the percent of students tested for the API, is expected to increase Hispanic API scores by 6.92 percent. However, as previously mentioned in the discussion of expected effects, increasing the number of students tested does not necessarily lead to higher API scores. Conversely, which students are tested may improved API scores, but improvements made in relation to this variable may be viewed as a strategy by a high school to unethically raise API scores by testing only those students that can perform well on tests. When further examining the predicted



coefficient interval ranges, the confidence intervals for percent of full credentialed teachers and percent of students tested are the broadest in comparison to average class size in core academic courses and percent Hispanic students. This suggests that the actual outcomes from altering these variables are likely to greatly vary. Thus, when implementing new policies with the goal of improving Hispanic API scores, it is important to understand the degree to which the results are likely to occur.

### **Evaluating the Research Question and Policy Implications**

The purpose of this regression analysis was to determine specific factors that influence changes in Hispanic API scores in California public high schools. The study's design examined various high school, student, and social factors that may be significantly correlated to Hispanic academic performance. Significant findings may increase discussions of Hispanic students' successes and failures in California's public education system. The results are disappointing in terms of magnitudes, as well as odd in regards to the results aligning with the expected directions. Perhaps a more narrowly focused study of a particular geographic region, school district, or individual high school may yield more significant results regardless of whether the outcomes coincide with previous research results.

Nevertheless, it is important to recognize that various factors are found to have statistically significant relationships to Hispanic API scores at confidence levels between 90 and 99 percent. Previous studies also support the fact that different inputs influence academic performance. Therefore, the more positive factors Hispanic students are exhibited to, the increased probability that these students will have higher API scores. Conversely, the more negative factors present, the increased likelihood that Hispanic students will have lower API scores. Consequently, policy-makers and educational administrators must consider multi-faceted solutions in creating greater success in the Hispanic student population. Based on this study, as well as previous findings, solutions to certain factors may be limited as the "issues" are out of the hands of public intervention methods.

### **Future Research**

When conducting future research on Hispanic API scores in various California high schools, studies may choose to examine individual counties in further detail. A general study may be conducted to examine regions of interest and draw comparisons by grouping the 58 counties in California into geographical areas, such as Northern, Central, and Southern California. However, aggregating county dummy variables may decrease the effects of the individual counties and further limiting the significance of the study. Nevertheless, it is suggested that county differences be further assessed. In addition, various variables may be of interest when

further studying Hispanic API scores. Factors, such as curriculum inputs, school district size, and school personnel composition, particularly in regards to ethnicity, are excluded in this particular study, but may impact Hispanic academic performance.

The preliminary regression on the various factors influencing Hispanic API scores is extremely broad, which may have adversely affected the results of the analysis. Granted, the factors that contribute to academic performance are difficult to understand as the majority of the findings may not be generalized across sample populations. This study provides limited information on the factors impacting Hispanic API scores. Thus, further research that is more thorough, perhaps examining more specific explanatory variables is warranted. A detailed examination and comparison of Hispanic API scores for high schools within a particular school district may prove to be an interesting Master's thesis.

### **REFERENCES**

- Baker, B., Keller-Wolff, C., & Wolf-Wendel, L. (2000). Two Steps Forward, One Step Back: Race/Ethnicity and Student Achievement in Education Policy Research [Electronic Version]. *Educational Policy*, 14(4), 511-529.
- California Department of Education. (2008). Executive Summary: A Framework for Closing California's Academic Achievement Gap [Electronic Version]. Retrieved on May 10, 2009, from <http://www.cde.ca.gov/eo/in/pc/documents/yro8es0122.pdf>
- California Department of Education (2008, November). 2008 Adequate Yearly Progress Report: Information Guide [Electronic Version]. Retrieved on May 10, 2009, from <http://www.cde.gov/ta/ac/ay/documents/infoguide08r.pdf>
- California Department of Education. (2009). Academic Performance Index (API). Retrieved on March 16, 2009, from <http://www.cde.ca.gov/ta/ac/ap/>
- California Department of Education. (2009). Achievement Gap Fact Sheet. Retrieved on May 10, 2009, from <http://www.cde.ca.gov/eo/in/se/agfactsheet.asp>
- Cheng, J.Y. (2001, November). "At Home and in School: Racial and Ethnic Gaps in Educational Preparedness." *Public Policy Institute in California: California Counts*, 3(2) [Electronic Version]. Retrieved on May 10, 2009, from [http://www.ppic.org/content/pubs/cacounts/CC\\_1101JCCC.pdf](http://www.ppic.org/content/pubs/cacounts/CC_1101JCCC.pdf)
- Driscoll, D., Halcoussis, D., & Svorny, S. (2003). School District Size and Student Performance. *Economics of Education Review*, 22, 193-201.

Goe, L. (2002). Legislating Equity: The Distribution of Emergency Permit Teachers in California [Electronic Version]. *Education Policy Analysis Archives*, 10(42).

Goodman, G., & Young, I. (2006). The Value of Extracurricular Support in Increased Student Achievement: An Assessment of a Pupil Personnel Model Including School Counselors and School Psychologists Concerning Student Achievement as Measured by an Academic Performance Index [Electronic Version]. *Educational Research Quarterly*, 30(1), 3-13.

PPIC. (2008, September). Just the Facts: California's Future Population. Retrieved on May 10, 2009, from [http://www.ppic.org/content/pubs/jtf/JTF\\_FuturePopulationJTF.pdf](http://www.ppic.org/content/pubs/jtf/JTF_FuturePopulationJTF.pdf)

Powers, J. (2003). An Analysis of Performance-Based Accountability: Factors Shaping School Performance in Two Urban School Districts [Electronic Version]. *Educational Policy*, 17(5), 558- 585.

Slovacek, S., Kunnan, A., & Kim, H. (2002). California Charter Schools Serving Low-SES Students: An Analysis of the Academic Performance Index. Retrieved on March 16, 2009, from Education Resources Information Center (ERIC), [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/1a/72/c7.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a/72/c7.pdf)

Studenmund, A.H. (2006). *Using Econometrics: A Practical Guide* [Fifth Edition]. New York, NY: Pearson Education, Inc.

Sung, H., Padilla, A., & Silva, D. (2006). Foreign Language Education, Academic Performance, and Socioeconomic Status: A Study of California Schools [Electronic Version]. *Foreign Language Annals*, 39(1), 115-130.

## Appendix A

## Study Methods and Research Findings on Affecting Academic Performance Index (API)

Publication Date, Authors	Location, N (Sample Size) Years of Data	School Level/Population Examined in Sample
(2000) Baker, et. al.	Nationally, 14,596 students, 1988 (longitudinal data)	Eighth grade students
(2003) Driscoll, et. al.	California, 5525 public schools representing 755 districts, 1999 Academic Performance Index scores	Elementary Schools, Middle Schools, and High Schools
(2002) Goe	California, 6387 public schools, 1999- 2000 Academic Performance Index scores	Various
(2006) Goodman and Young	Pacific coast state, 150 public school districts, API year unspecified	N/A – school districts
(2003) Powers	Los Angeles Unified School District, San Diego Unified School District, 534 public schools, comparison of 1999 and 2001 API scores	Elementary schools
(2002) Slovacek, et. al.	California, all California public schools, comparison of 1999, 2000, and 2001 API scores	N/A – Charter vs. Non-Charter Schools (with a focus on low socioeconomic schools)
(2006) Sung, et. al.	California, 200 public schools, 2003 API scores	High Schools

## Appendix A

## Study Methods and Research Findings on Affecting Academic Performance Index (API)

Publication Date, Authors	Details Provided on the Methods of Statistical Analysis	School Level/Population Examined in Sample
(2000) Baker, et. al.	Mean comparison, Ordinary Least Squares Regression	Regression on aggregate race groups indicate that whites and Asians outperform other groups. Under the Hispanic subgroup, Cuban students perform the best, while in the Asian classification, it is dependent on the subject as to which ethnicity performs better. Socioeconomic status and language proficiency play significant roles.
(2003) Driscoll, et. al.	Regression analysis (for each school level)	District size is negatively correlated with API scores at the elementary and middle school level. No statistical significance found for high schools.
(2002) Goe	Multiple regression analysis	Found multiple student characteristics variables to be statistically significant and negatively correlated with API scores (i.e. socioeconomic status, parents' education, % of emergency permit teachers).
(2006) Goodman and Young	Regression analysis	The number of school psychologists has a significant positive correlation with school district's API scores.
(2003) Powers	Regression analysis	The districts differed in the sign of the coefficient for teachers' possessing full credentials. However, in both districts' traditional calendar schools and schools with students of a higher socioeconomic background have higher API scores.
(2002) Slovacek, et. al.	Correlation analysis, regression analysis	Charter schools serving high populations of low socioeconomic status students outperform non-charter schools serving the same students in API scores and are more likely to improve annually.
(2006) Sung, et. al.	Regression analysis, descriptive statistics and T-tests	A statistically significant positive correlation is found between enrollment in foreign language classes and school API scores.

## Appendix A

## Study Methods and Research Findings on Affecting Academic Performance Index (API)

Publication Date, Authors	School Level/Population Examined in Sample
(2000) Baker, et. al.	Math performance: results indicate significant performance percentage decreases in scores for blacks (6.78), Native Americans (7.10), and Hispanics (3.85). Reading performance: a 3.86
(2003) Driscoll, et. al.	For a 1% increase in district size, there is a 5.27% decrease in base API scores for elementary schools, a 4.00% decrease in base API scores for middle schools, and a 1.42% decrease in base API scores for high schools.
(2002) Goe	A 1% increase in students qualifying for free or reduced price lunches is associated with a 1.47% decrease in API scores. A 1% increase in the percent of Hispanic students results in a 9.1% decrease in API scores. A 1% increase in percent of parents without a high school diploma results in a 1.18% decrease in API scores. Lastly, a 1% increase in the percent of emergency permit teachers is associated with a .62% decrease in API scores. Conversely, a 1% increase in the percentage of parents that attended graduate school is associated with a 2.01% increase in API scores.
(2006) Goodman and Young	A one unit increase in the number of school psychologists employed in the school district results in a statistically significant .69% increase in base API scores, while a one unit increase in the number of school counselors employed in the school district only results in a .18% increase in API scores, but this effect is found to be insignificant.
(2003) Powers	(2001 results) Los Angeles Unified: a 1% increase in percentage of students in reduced price meal plans and percent of ESL students results in decrease of 2.52% and .30% in base API scores respectively. San Diego Unified: 1% increase of students in meal plans and percent of ESL students yields a 2.52% and 1.06% decrease in API scores.
(2002) Slovacek, et. al.	1% increase in the percentage of ESL students results in a .56% decrease in API scores. A 1% increase in percentage of students participating in lunch programs is associated with a 2.65% decrease in API scores. Conversely, a 1% increase in the percentage of teachers with full credentials results in a 1.06% increase in base API scores.
(2006) Sung, et. al.	Specific magnitudes unreported. However, researchers conclude that as the percent of students enrolled in foreign language classes increases, API scores increase.

Appendix B

Interactive Scatter Plot: Test for Heteroskedasticity

