RESEARCH REPORT 2011-2

CollegeBoard
inspiring minds™

Investigating Grade Inflation and Non-Equivalence

By Kelly E. Godfrey

Kelly Godfrey is an associate research scientist at the College Board.

Acknowledgments: The author would like to thank Pamela Kaliski, assistant research scientist, and Rosemary Reshetar, senior psychometrician in Research & Development for their review of and helpful comments on earlier drafts of this paper.

GRADIN

Contents

Executive Summary 2

Introduction 3

Methods 5

Study 1: Grade Inflation 5

Study 2: Grade Non-Equivalence........ 5

Results 6

Study 1: Grade Inflation 6

Study 2: Grade Non-Equivalence........ 9

Discussion 14

Conclusion 15

References........ 16

Figures

Figure 1: Rise in Overall High School GPA for High School Diploma Recipients 7

Figure 2: SAT–M Scores for Classes of 1996 and 2006 7

Figure 3: SAT–V Scores for Classes of 1996 and 2006 8

Tables

Table 1: Character to Numeric Grade Values 6

Table 2: Regression of Cumulative High School GPA on Gender and Ethnicity........ 8

Table 3: Correlation and Regression Coefficients: Biology........ 10

Table 4: Correlation and Regression Coefficients: Calculus AB........ 11

Table 5: Correlation and Regression Coefficients: English Language 12

Table 6: Correlation and Regression Coefficients: English Literature........ 13

Table 7: Correlation and Regression Coefficients: U.S. History 13

Executive Summary

When educators see that two students received a B in a particular course in two different schools and/or years, does it indicate the same level of understanding and achievement for both students? This paper addresses the assumption that grades are equivalent within subjects across schools and years. By comparing course grades to a standardized test score that is comparable across all schools, districts, and states, this relationship is investigated. Data used in this study include public high school student records from one diverse state in the continental U.S., along with corresponding exam score records from the College Board. To explore grade inflation, where grades increase without corresponding increases in achievement, cumulative high school grade point averages are measured across time and compared to changes in the math (SAT-M) and verbal SAT-V) sections of the SAT®. To measure grade non-equivalence across schools, five Advanced Placement Program® (AP®) subjects were chosen for comparison across five large high schools. The AP Exams serve as a standard measure of skills and knowledge obtained by the student. They are compared to AP course grades, which are assigned by individual classroom teachers and based on a variety of criteria at the teachers' and schools' (and possibly districts') discretion. Results indicate a presence of grade inflation across years and non-equivalence across schools.

Introduction

As academic rigor and expectations of college-bound high school students rise, the subject of grade inflation and non-equivalence is of growing concern for universities and college admission councils. Exam scores such as the SAT and AP, standard across schools and districts for each subject or content area, are one way to discern differences in skills and knowledge between students from different classrooms and grading standards. Although course grades are intended to indicate some measure of proficiency, some teachers report that they are encouraged to keep grades high regardless of proficiency, especially for high-achieving students, a practice that not only leads to inflation and lowered variability but also makes decisions more difficult for college admission counselors.

Grade inflation, the phenomenon of rising grades despite a lack of corresponding increase in achievement, is not new to higher education or to the College Board. However, the volume of academic literature as it pertains to secondary education is a bit weak, with a majority of the research devoted to grade inflation within a higher education setting. In 1997, Ziomek and Svec investigated the phenomenon of rising high school grades. The authors claimed that it is impossible to tell if grade inflation is a persistent problem without several years of data. Using ACT's student history files from 1989 to 1994, Ziomek and Svec studied the existence, persistence, and degree of grade inflation at 5,136 college-preparatory high schools in the United States. Examining self-reported data, they calculated an overall grade point average in four core course areas: mathematics, science, social sciences, and English. Dividing the sample into deciles according to composite ACT scores, they noted a steady increase in grade point averages within each decile.

Although course grades are intended to indicate some measure of proficiency, some teachers report that they are encouraged to keep grades high regardless of proficiency.

In 2003, the College Board published a research report by Camara, Kimmel, Scheuneman, and Sawtell that gives an extended history of the grade inflation debate among academics and educators and addressed the notion of grade inflation over a 25-year period. Camara and colleagues argue that teachers would be more likely to base grades solely on achievement if they had a guarantee that those grades would be used only for denoting achievement. Because that promise is not possible, teachers use more than just achievement on which to base their assigned grades. Using self-reported grade point averages on the Student Descriptive Questionnaire (SDQ) and scores on the College Board's SAT, researchers at the College Board randomly sampled from eight cohorts of graduating seniors between the years of 1976 and 2002 to study the increase in self-reported grades over time. Conducting multiple regression using gender, ethnicity, and parental education level to predict reported grade point average for each cohort, Camara and colleagues found that despite little to no increase in SAT-V and SAT-M scores, the grade point average of the college-bound student has increased over the last several decades. The authors also found evidence of a slightly wider gap between students whose parents have higher levels of education, as well as differences in grade inflation across race/ethnicity and gender groups, with a large portion of the increase occurring in the last years of the study, from 1998 to 2002.

Grade inflation is not just a concern in the United States. In the mid-1990s, Sweden's educational system changed from a norm-referenced, centralized model to criterion-referenced decentralized model. Students had more autonomy in choosing the classes they took, and grades were determined solely through classroom assessment by the teachers. Educational researchers noted increases in grades as well as the dropout rate of students in upper secondary schools. Using the SweSAT, Sweden's equivalent to the SAT, researchers determined that high school grades have indeed increased, on average, from 1997 to 2002 while standardized test scores have not (Wikström, 2005). Using multiple regression, the author rules out the notion that grades are increasing due to student selection effects or making easier course choices and presents one possible remaining explanation: lowering of grading standards.

Also familiar to many educational researchers is the notion of differential grading standards for the same curriculum and material, noted here as "grade non-equivalence." Beginning in fall of 2005, the public university system in the state of Oregon planned to begin fully adopting a proficiency-based system for admission standards. David Conley's paper presentation at the 2000 annual meeting of the American Educational Research Association focused on the baseline "piloting" done by 50 high schools in the state in the late 1990s. He found that proficiency scores assigned by trained raters and grades assigned by classroom teachers have low correlations, suggesting different constructs being measured by each. Conley notes that proficiency scores were assigned using written assignments, unit tests, and projects, while grades are often based on information beyond these three criteria. Conley also noted that no two teachers in his study had the same grading system, resulting in students taking the same-named course with different grading criteria.

In 2004, researchers at ACT released two research reports exploring and documenting the issues of grade inflation and differing grading standards among high schools. To investigate differential grading standards, Woodruff and Ziomek (2004b) used high schools' average ACT scores to divide schools into quintiles for each year: 1998 to 2002. Then, using the first (schools with lowest average ACT scores) and fifth quintile (schools with highest average ACT scores), they calculated correlations between ACT scores and high school GPAs, and regressions of high school GPAs on ACT scores across all five years of the study. They concluded that different high schools employ different grading standards, and that a student's grades depend not only on his own achievements but on the achievements of his schoolmates as well. For the study of grade inflation, Woodruff and Ziomek (2004a) compared self-reported high school GPAs to ACT scores across 13 years: 1991 to 2003. They found that grades increased over time without a concurrent increase in achievement, as measured by the ACT, thus providing general evidence of grade inflation over the years studied.

Researchers have presented fairly strong evidence that grade inflation, as well as differential standards, are a problem in education. Using self-reported information and results from some sort of standardized measure of ability or proficiency, they have shown that grades have been increasing over the years with no corresponding increase in proficiency scores and that teachers and schools do have an effect on the relationship between grades and true ability. This research study explores the presence and severity of grade inflation and differential grading standards (i.e., non-equivalence) across schools within the same subject area using actual grades and achievements on AP Exams and the SAT.

Methods

The data used in this study are student-level course grades, overall GPAs, and 1996–2006 exam scores from one large, diverse state public school system. These student records were then matched to data from the College Board, including SAT scores and AP Exam scores. This paper presents two key investigations: grade inflation and grade non-equivalence. The first study uses SAT scores and cumulative high school grade point averages across 11 years of diploma-receiving cohorts and explores the possible presence of grade inflation. The second uses AP course grades and exam scores to investigate differential grading standards in several large high schools.

This paper presents two key investigations: grade inflation and grade non-equivalence.

Study 1: Grade Inflation

Final cumulative high school GPAs for all high school diploma recipients were compared from 1996 to 2006 to note general trends in GPAs over time. The total number of graduates across the 11 cohorts is approximately 1.2 million students. To measure the overall change in high school GPA across the time span in comparison to SAT scores, just the class of 1996, class of 2001, and class of 2006 students were used. For SAT scores, only SAT-M (math) and SAT-V (verbal) scores were used because the writing portion of the SAT was not added until March 2005. An SAT-M and SAT-V composite score is avoided here because the two sections measure different constructs. The author believes some value would be lost by combining the scores into one variable.

In order to better understand the degree or magnitude of inflation for different gender and race/ethnic groups of students across three graduating cohorts, a series of three multiple regressions were calculated, similar to Camara et al. (2003), one for each cohort. The dependent variable was the reported cumulative final high school GPA for students receiving a high school diploma in one of the three years of interest. The independent variables were gender (0 = male, 1 = female) and dummy coded variables representing race/ethnicity.

Study 2: Grade Non-Equivalence

To investigate the presence and magnitude, if applicable, of grade non-equivalence across schools, five AP courses were chosen due to their relatively large popularity and differences in content: Biology, Calculus AB, English Literature, English Language, and U.S. History. U.S. History and English Language are taken primarily by 11th-graders. For each subject, the five high schools with the largest number of students taking both the course and the exam in the 2004-2005 academic year were compared.

Advanced Placement® courses offer a unique opportunity to investigate grade non-equivalence between high schools due to the uniform exam offered to all participating students. This allows researchers to compare not only student proficiency in each school but also the relationship between course grades using the exam scores as a uniform measure of aptitude in the content area. Final course grades were not uniformly available for every student because schools use different scheduling and grading systems. Grades could be assigned yearly, semesterly, trimesterly, quarterly, per six-week period, and so on. In addition, some schools offer AP courses taught over one term (semester, trimester, etc.) while others make them into yearlong curricula. Because of these discrepancies, mean grades were calculated for each student, and these values serve as final course grades. Grades were originally available as characters and were recoded as numeric values. Table 1 shows the recoded values.

Table 1

Character to Numeric Grade Values

Grade	Value	Grade	Value	Grade	Value
A+	4.0	**C+**	2.3	**F**	0.0
A	4.0	**C**	2.0		
A-	3.7	**C-**	1.7		
B+	3.3	**D+**	1.3		
B	3.0	**D**	1.0		
B-	2.7	**D-**	0.7		

For each high school, a correlation between final course grades and AP Exam scores was calculated. AP Exam scores were then regressed onto AP course grades and school to better understand the differences in the relationship between proficiency scores and assigned course grades across school settings. Results are presented in the following section.

Despite the general rise in grades, standardized scores on the SAT remained relatively unchanged.

Results

Study 1: Grade Inflation

The average GPA for the class of 1996 was 2.64 on a scale of 0 to 4.0. However, in 2006 the average GPA was 2.90, an increase of 0.26, or just over a quarter of a grade point, with a 47.5 percent increase in the number of graduates, from n = 87,721 to n = 129,428. Figure 1 shows the general trend of rising overall grade point averages across the 11 cohorts included in this study. Despite the general rise in grades, standardized scores on the SAT remained relatively unchanged. The distributions of SAT-M and SAT-V scores for the class of 1996, class of 2001, and class of 2006 are presented in Figures 2 and 3. In 1996, the average SAT-V score for a graduating senior was 497.06, and for SAT-M, it was 497.12. In 2006, the average SAT-V score was 495.15 for graduating seniors and the mean SAT-M score was 497.83.

Figure 1
Rise in Overall High School GPA for High School Diploma Recipients

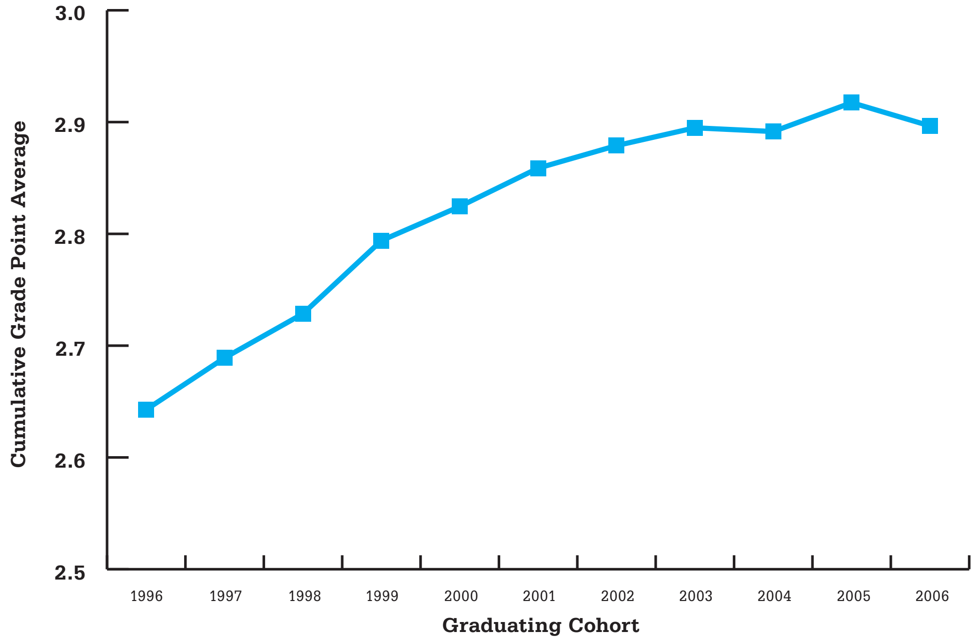


Figure 2
SAT-M Scores for Classes of 1996 and 2006

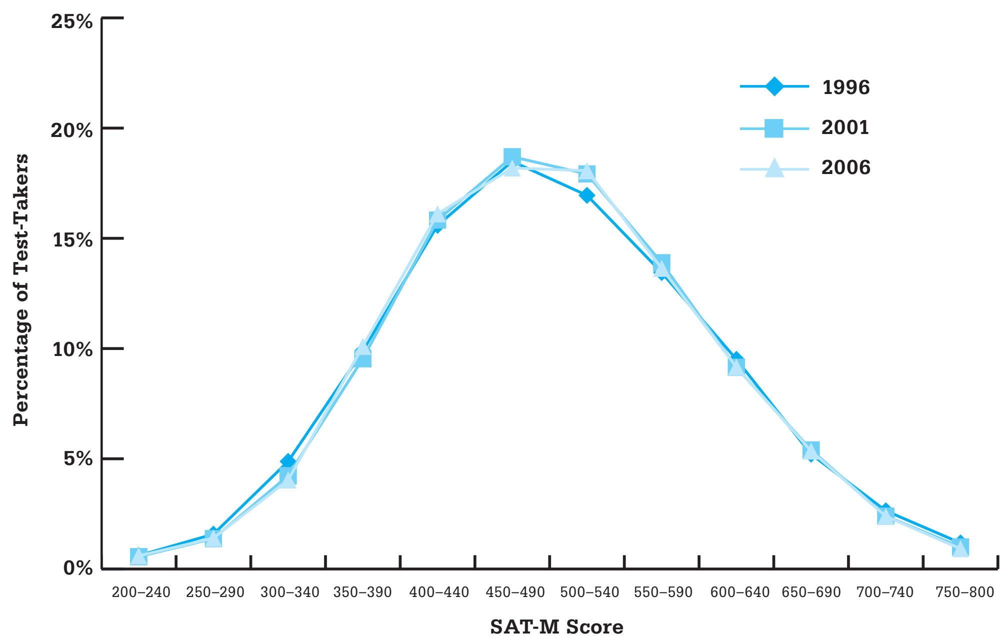
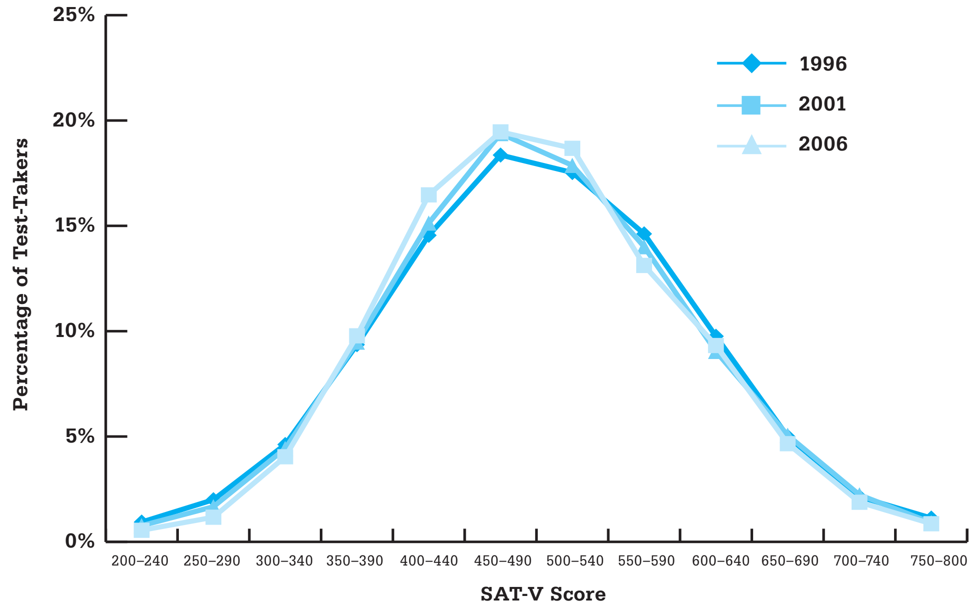


Figure 3
SAT-V Scores for Classes of 1996 and 2006



In keeping with previous research regarding the phenomenon of grade inflation, three multiple regressions were computed to understand the inflation differences among different gender and racial/ethnic groups. *R*-squared values, or percentage of variance explained, unstandardized beta coefficients, and semipartial correlations are presented in Table 2. One value to note is the increasing constant value, or intercept, across the three cohorts presented here: from 2.636 in 1996 to 2.884 in 2006. This indicates that the expected values of grade point averages, after taking gender and ethnicity in account, are rising over the years. Gender and all race/ethnic variables were statistically significant beyond the 0.01 level except for Race = Other (p = 0.237), which may have been due to small n size in that category.

Table 2
Regression of Cumulative High School GPA on Gender and Ethnicity

	1996		2001		2006	
Variance Explained	9.8%		10.2%		9.6%	
Coefficients:						
Constant	2.63		2.86		2.88	
Gender	0.22		0.19		0.22	
Asian	0.24	n = 2,414	0.16	n = 3,015	0.18	n = 3,978
Black	-0.39	n = 18,336	-0.35	n = 21,736	-0.36	n = 24,166
Hispanic	-0.24	n = 12,982	-0.24	n = 17,372	-0.20	n = 25,281
Other Ethnic	-0.04*	n = 191	-0.04	n = 854	-0.07	n = 2,275
Semipartial correlations:						
Gender	0.17		0.17		0.18	
Asian	0.06		0.05		0.05	
Black	-0.24		-0.24		-0.23	
Hispanic	-0.13		-0.15		-0.13	
Other Ethnic	0.00		-0.01		-0.02	

* denotes p>0.05

Gender was coded 0 = male, 1 = female, so cumulative high school GPAs are higher in all three cohorts for females than for males. Removing the effects for ethnicity, females are 0.225 grade points higher than males in 1996, and 0.221 grade points higher in 2006. These findings suggest that although males and females have different mean GPAs, this gap has not overly expanded or condensed over the decade observed here.

In all three observed cohorts, after removing gender effects, black, Hispanic, and other students identified as other races/ethnicities were at a disadvantage with overall high school GPA, with black students tending to achieve the lowest grades. Asian students, on the other hand, have shown higher grade point averages across all three cohorts. While these overall differences among race/ethnic groups have decreased from 1996 to 2006, the reductions have not been impressive and do not appear particularly meaningful.

Although males and females have different mean GPAs, this gap has not overly expanded or condensed over the decade observed here.

By squaring the semipartial correlations, the contribution of each predictor to the overall variance of GPA can be determined. Therefore, gender of the student and whether or not he or she is black or African American contribute the most of the included predictors to the variance of GPA: approximately 2.9 percent and 5.8 percent, respectively. The contributions of the independent variables to the overall variance of the dependent variable (GPA) appear stable across the three cohorts, indicating that the relationship among gender and ethnicity and high school GPA does not change meaningfully over the time period explored here.

Study 2: Grade Non-Equivalence

Correlations between AP course grades and AP Exam scores were calculated for the five largest high schools for each of the five chosen AP subjects and are presented in Tables 3 through 7. School names and ID numbers are removed and are represented with letters. Overall, 15 schools were observed, with one school (school F) included in all five subject analyses. It may be useful to note that AP Exam scores range from 1 to 5, and AP course grades range from 0.0 to 4.0. While a score of 3 on the AP Exam corresponds to a mid-level B to mid-level C college-level performance, a 3.0 AP course grade is a B.

In addition to calculating correlations, AP Exam scores were regressed onto AP course grades and school, which was dummy coded. School F was the only school that was included in each subject analysis, so in the dummy coded variables representing schools, school F was excluded. A zero in all school categories indicated that a student was enrolled in this school.

Table 3 presents the regression results for AP Biology. The correlations between exam score and course grade spanned a fairly wide range for this subject, ranging from 0.29 at school H to 0.77 at school F. The overall model explained 58.7 percent of the variance in AP Exam scores, the highest amount of the five subjects explored here. A student in school F with an A in AP Biology would likely obtain a 4 on the AP Exam (4.0 * (0.89) + 0.77 = 4.33). In this scenario, all four schools in the model have negative relationships with exam scores. School H shows the most negative coefficient, indicating that the grades assigned to students have a weaker positive relationship to AP Exam scores. For instance, students at School H, on average, receive a B+ or better in the course, but they tend to perform at an AP level of a D.

Final course grade explained 27 percent of the variance in AP Biology Exam scores, as shown by squaring the semipartial correlation. However, whether or not a student is enrolled at School H accounts for 32.5 percent of the variance in AP Biology Exam scores, and whether or not a student attends School O accounts for 19.4 percent of variance. These values indicate that there is a school effect, apart from course grades, that explains for variability in exam scores. These school effects are not uniform across schools.

Table 3

Correlation and Regression Coefficients: Biology

School	Pearson *r*	Mean Score	Mean Grade	*N*
A	0.60	3.31	3.49	105
F	0.77	3.23	2.75	167
H	0.29	1.52	3.69	94
K	0.64	2.76	3.00	139
O	0.52	1.81	2.92	158
Variance Explained	58.7%			
Weights:		**Semipartial Correlations:**		
Constant	0.77			
Grade	0.89		0.52	
School A	-0.58		-0.14	
School H	-2.55		-0.57	
School K	-0.69		-0.18	
School O	-1.57		-0.44	

Table 4 presents the regression results for the subject of AP Calculus AB. Correlations between course grades and exam scores spanned a tighter range than those of AP Biology: from 0.45 to 0.63. After the effects of schools B, G, J, and K are taken into account, a student who received an A in the course is expected to achieve about a 4 on the exam. Schools B, G, and K showed negative beta coefficients, while School J's coefficient was positive. School J also had the highest mean exam score among the schools and one of the lowest mean course grades, providing evidence of differential grading standards compared to other schools.

These values indicate that there is a school effect, apart from course grades, that explains for variability in exam scores.

Course grade accounted for approximately 29.2 percent of variance in AP Calculus AB Exam scores. Whether or not a student attended School G accounted for 3.2 percent of variance, which is the highest squared semipartial correlation among all school indicators in the model. Unlike AP Biology, Calculus AB has much weaker school effects that are more similar to each other, although not identical. School effects do not appear to be noteworthy in this model.

Table 4

Correlation and Regression Coefficients: Calculus AB

School	Pearson *r*	Mean Score	Mean Grade	*N*
B	0.63	2.80	2.98	100
F	0.57	2.68	2.64	206
G	0.51	2.65	3.49	92
J	0.63	3.21	2.71	92
K	0.45	2.72	3.20	92
Variance Explained	30.8%			
Weights:		**Semipartial Correlations:**		
Constant	0.21			
Grade	0.94		0.54	
School B	-0.20*		-0.05	
School G	-0.83		-0.18	
School J	0.46		0.10	
School K	-0.49		-0.11	

* denotes p>0.05

Table 5 presents results from the regression using AP English Language. Correlations between grades and exam scores tended to be on the lower end when compared to some other subjects, ranging from 0.35 to 0.60, and the regression model only explained 33.4 percent of the variance in AP Exam scores. A student at School F with an A for his course grade is expected to achieve at least a score of 4 on the AP Exam. For this subject, all four schools in the model present negative beta coefficients, indicating some degree of disparity between grades and proficiency scores, and a small degree of differential grading standards from each other.

For AP English Language, final course grade explains approximately 18.5 percent of the variance in AP Exam scores. School indicator variables also contribute to the variance: from 5.3 to 10.9 percent. The contributions of these school indicator variables to explaining the variance in AP Exam scores does not appear particularly meaningful here.

Four schools in the model present negative beta coefficients, indicating some degree of disparity between grades and proficiency scores.

Table 5

Correlation and Regression Coefficients: English Language

School	Pearson *r*	Mean Score	Mean Grade	*N*
C	0.60	2.34	2.78	209
F	0.56	3.44	3.01	232
H	0.35	2.76	3.58	193
I	0.36	2.63	3.34	196
M	0.44	2.89	3.29	187
Variance Explained	33.4%			
Weights:		**Semipartial Correlations:**		
Constant	1.52			
Grade	0.64		0.43	
School C	-0.95		-0.31	
School H	-1.04		-0.33	
School I	-1.02		-0.33	
School M	-0.73		-0.23	

Table 6 presents the regression results for the subject of AP English Literature, the largest AP course and exam, with one school reporting over 300 students enrolled in the course and taking the exam in one year. Correlations tended to be somewhat low, ranging from 0.23 to 0.56, with only 26.2 percent of the variance in AP Exam scores explained by the model, which is the lowest amount of the five subjects. Only schools K, N, and O presented negative beta coefficients, indicating some degree of higher grades awarded to lower levels of proficiency than schools C and F. At school F, a student who receives an A for the course is expected to achieve either a 3 or a 4 on the AP Exam. Note that average grades for this subject are relatively high, with four out of five schools awarding above a B (3.0), on average. The difference in beta weights also provides evidence that different schools have different grading standards, or that a student's school has an effect on the student's relationship between his course grade and his exam score.

Final course grades explain approximately 20.3 percent of the variance in AP English Literature Exam scores. School enrollment variables explain from 0 to 5.8 percent of the variance, which is not particularly meaningful here. Whether or not a student attends school C does not contribute to explaining the variance between AP English Literature Exam scores, once the other school indicators have been taken into account.

Table 6
Correlation and Regression Coefficients: English Literature

School	Pearson r	Mean Score	Mean Grade	N
C	0.41	2.85	2.77	205
F	0.56	3.07	3.13	303
K	0.23	2.95	3.26	213
N	0.39	2.50	3.41	195
O	0.55	2.49	3.04	297
Variance Explained	26.2%			
Weights:		**Semipartial Correlations:**		
Constant	1.13			
Grade	0.62		0.45	
School C	0.00*		0.00	
School K	-0.20		-0.07	
School N	-0.75		-0.24	
School O	-0.52		-0.19	

* denotes p>0.05

Table 7 presents the regression results for AP U.S. History. Correlations between grades and exam scores are the highest in this subject when compared to the other four presented here, ranging from 0.65 to 0.77. Consequently, the regression model explained the 51.5 percent of the variance in exam score performance. All four schools in the model presented negative beta coefficients, indicating again some degree of higher grades being awarded despite lower proficiency.

Final AP U.S. History grades explain approximately 43.6 percent of the variance in AP Exam scores. Whether or not a student is enrolled in school D or L accounts for 7.8 or 5.8 percent, respectively. Because the beta coefficients range from -0.37 to -1.07, there is some indication that schools are using different grading standards for what should be the same proficiency. For example, schools A and E account for 2.9 and 1.0 percent each. Students at school A had the highest course grade on average (B+). They also had the highest mean AP U.S. History Exam score (3.29), while students at school F earned a similar average exam score (3.18) but tended to receive lower course grades (C+ to B-).

Table 7
Correlation and Regression Coefficients: U.S. History

School	Pearson r	Mean Score	Mean Grade	N
A	0.77	3.29	3.39	171
D	0.70	2.57	2.99	235
E	0.65	2.36	2.09	270
F	0.73	3.18	2.54	290
L	0.68	2.83	3.19	181
Variance Explained	51.5%			
Weights:		**Semipartial Correlations:**		
Constant	0.61			
Grade	1.01		0.66	
School A	-0.75		-0.17	
School D	-1.07		-0.28	
School E	-0.37		-0.10	
School L	-1.00		-0.24	

Discussion

Results presented here should come as little or no surprise to most educators. Previous research on the grading practices of teachers has found that a majority of teachers use constructs beyond pure achievement to grade their students. These include, but are not limited to, effort and cooperation (see Camara et al, 2003 for brief overview of the literature). The notion that teachers use their own mechanisms for assigning grades, sometimes defying already well-defined protocols, is not new to educational researchers (McMillan, 2001). Unfortunately, many research attempts to understand this notion have been based on small convenience samples, or have been forced to use achievement or aptitude scores that may be a bit more removed from the curriculum on which the grades are based than what would be ideal.

Regression analyses in all five subjects showed differential grading practices between schools, with some offering higher grades to lower levels of proficiency than the other schools.

A major advantage of using Advanced Placement scores is that the exams, by definition, cover the curriculum and provide a proficiency level used by high schools and colleges alike. Unfortunately, AP courses tend to be taken by higher-achieving students (within the school), often not including students from across the range of ability levels, thus limiting the generalizations that can be made regarding secondary education in the United States.

Calculus AB correlations did not span as wide of a range as the other subjects. This may be due to the hierarchical nature of mathematics. Calculus is also not a required math course in high school curricula, so results may differ from other subjects where students are required to take either an AP or non-AP version of the course. Previous research has indicated more agreement between teachers and proficiency raters in the area of mathematics than other content areas (Conley, 2000). However, regression analyses in all five subjects showed differential grading practices between schools, with some offering higher grades to lower levels of proficiency than the other schools. In an ideal setting, a researcher would isolate the classroom variance, gaining a better understanding of the different grading standards between teachers. However, due to the nature of the data, this information is unavailable, and only the effects of the department can be isolated.

Conclusion

If researchers and postsecondary institutions' admission officers understand that schools (and even teachers) have dissimilar grading standards, how are students compared in high-stakes situations, such as admission decisions? Some research promotes the notion of rank-ordering students to better understand their level of achievement in reference to their peers. However, this does not appease the concern that grading is subjective, and a student's grade reflects not only his achievement but the achievement of his classmates as well. When comparing schools, it is not uncommon to see that despite seemingly equal grades, scores on achievement tests show great differences in the student populations (Jost, 2002).

When comparing schools, it is not uncommon to see that despite seemingly equal grades, scores on achievement tests show great differences in the student populations.

This phenomenon happens not only across classrooms and schools but also over years with new cohorts of students. It is not confined to secondary education. Colleges and universities have focused on the growing problem as well, noting that students often choose courses and majors based on the grading ease of the professor or department. They are beginning to avoid more technical courses such as mathematics and science due to perceived harder standards (Johnson, 1997). With the growing competition for spots in higher educational institutions, internships, post-baccalaureate opportunities, and even jobs, everyone is aware that grades play an important role. Teachers are forced to choose between adhering to grading guidelines they may deem inappropriate and using grades as motivators to reward good efforts rather than achievement. Evidence suggests that many teachers choose the latter. With the problem of grade inflation and non-equivalence, decisions meant to be based on achievement may not be accurate, and comparisons made across groups may be faulty as well.

References

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E. A. (2003). Whose grades are inflated? (College Board Research Report No. 2003-4). New York, NY: College Board.

Conley, D. T. (2000). *Who is proficient? The relationship between proficiency scores and grades*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Johnson, V. E. (1997). An alternative to traditional GPA for evaluating student performance. *Statistical Science*, 12, 251–278.

Jost, K. "Grade Inflation." *CQ Researcher* June 7, 2002: 505–520. *The CQ Researcher Online*. CQ Press. www.cqpress.com.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practices*, 20(1), 20–32.

Wikström, C. (2005). Grade stability in a criterion-referenced grading system: the Swedish example. *Assessment in Education*, 12(2), 125–144.

Woodruff, D. J., & Ziomek, R. L. (2004a). *High school grade inflation from 1991 to 2003*. (ACT Research Report Series 2004-4). Iowa City, IA: ACT.

Woodruff, D. J., & Ziomek, R. L. (2004b). *Differential grading standards among high schools*. (ACT Research Report Series 2004-2). Iowa City, IA: ACT.

Ziomek, R. L., & Svec, J. C. (1997). High school grades and achievement: Evidence of grade inflation. *NASSP Bulletin*, 81, 105–113.

CollegeBoard
inspiring minds™

11b-3181