CollegeBoard

Getting to Know Your Criterion: Examining College Course Grades and GPAs Over Time

Jessica Marini, associate research scientist at the College Board

Emily Shaw, senior director at the College Board

Linda Young, senior database statistician

Maureen Ewing, senior director at the College Board

Executive Summary

This study investigated differences in college grading practices (first-year grade point average and course grades) by student and institutional characteristics and by academic discipline to inform and improve our understanding and use as among the most commonly employed criteria in validity and college readiness research. In addition, trends in college grades were examined over a five-year period to determine the stability of these grading differences. Findings show that there were small increases in overall FYGPA from 2007 to 2011 and that FYGPA tended to vary by student and institutional subgroups. There were also major differences in average course grade by academic discipline and these differences remained after controlling for students' SAT® scores and institutional selectivity. These findings can serve to contextualize higher education research studies that include college grades as predictors and/or outcomes and can ultimately inform and impact key educational benchmarking and policy work at local and national levels.

Contents

Tables

Figures

Introduction

In the higher education realm, there has always been a struggle to define the meaning(s) of success along with its measurement—for research purposes as well as for accountability purposes. For example, is degree completion most appropriate for capturing college success, or should the focus be on performance in coursework, or perhaps on employment or earnings after graduation? It is likely that all measures (including others not named), and/or a combination thereof, are important to different constituencies. Yet, for traditional college admission research, the most commonly studied outcomes include freshman and cumulative grade point average (GPA) (Camara & Echternacht, 2000).

Camara (2005) outlines a number of reasons why the GPA is such a widely utilized criterion, including that it is among the most commonly available measures in higher education, it is highly related to available and important academic predictors such as high school grades and standardized test scores, and the GPA is often considered a proxy of sorts for more long-term criteria such as graduation, employment, or success in graduate school. But we know that the GPA as a criterion is not without its pitfalls. Such issues may include criterion unreliability, grade inflation, differences in course-taking patterns across students, and restriction of range in the assignment of grades, among others (Camara & Quenemoen, 2012).

Ramist, Lewis, and McCamley (1990) extensively studied the use of the first-year GPA (FYGPA) as a criterion in SAT® predictive validity research across 38 colleges in the 1980s. The study was undertaken because, as they noted, "Even though unrelated to the true validity of the predictor, any limitation in the criterion reliability or comparability, or restriction of range in predictor scores, reduces the observed correlation between the predictor and the criterion" (pp. 253–254). Their study focused on various aspects (e.g., student course load, the variety of courses taken, the grading leniency in particular courses, etc.) of the FYGPA to better understand how these factors might relate to observed correlations between the SAT and FYGPA. Among their more noteworthy conclusions, Ramist et al. found that there is a great deal of noncomparability in student FYGPA, with some substantial differences in the average grade assigned to a course and a student's predicted FYGPA (based on SAT scores and HSGPA). In particular, course grades were lowest in relation to the student's ability in the mathematics/science fields while students with lower SAT scores tended to take courses that were graded more leniently. Ramist et al. point out that the benefits in reliability of utilizing a FYGPA composed of multiple courses as the criterion in validity research (as opposed to a course grade) seem to be outweighed by the noncomparability of college course grades. They recommend that future validity research reports correlations that are derived from the relationship between the predictor and each course separately and then averaged across the appropriate courses within each student rather than a general FYGPA that is plagued with the aforementioned issues.

Building on the work of Ramist et al. (1990), Berry and Sackett (2009) further examined the utility of the academic performance composite (APC), calculated across parallel individual course grades (controlling for student differences in course choice), as the criterion in admission validity research across students who entered 41 colleges in the mid-to-late 1990s. They compared this composite of parallel course grades to the traditional FYGPA and cumulative GPA (through completion) to determine whether the use of FYGPA or cumulative GPA as the criterion in SAT predictive validity research is underestimating the relationship between SAT scores and college outcomes due to individual differences in course selection. Berry and Sackett found that using the traditional FYGPA and cumulative GPA as the criteria underestimated the variance accounted for by SAT scores and HSGPA by about 30% to 40%. SAT validity estimates calculated at the individual college course level were consistently strong and positive across the colleges in the study and tended to be about .13 higher than the traditional GPA measures. These findings support those of Ramist et al. (1990), suggesting that the construct irrelevant variance in the FYGPA measure is impeding our understanding of the true relationship between SAT scores and academic performance in college.

The current study will explore the college-grading practices literature in order to better understand the information that might be included in the assignment of a grade, and will empirically examine the differences in grading practices by academic discipline and institutional characteristics, as well as any pertinent trends in college grades that researchers should be aware of. Such analyses can inform and improve our understanding and use of one of our most commonly utilized criteria in validity and college readiness research, which ultimately impacts educational benchmarking and critical policy work at local and national levels.

What’s in a Grade?

The intended purpose of a grade in a college course is to signify a student’s level of academic achievement in the domain over a course of study (Allen, 2005). Unlike standardized test scores where great care is taken to remove construct irrelevant variance or measurement error in the evaluation of student performance, college grades that are assigned by professors often include various types of information about the student—some closely tied to academic achievement and some less so. For example, grades may additionally reflect student attendance, participation, effort, conscientiousness, or even instructor biases, attitudes, expectancies, or particular goals of the instructor (Brookhart, 1993; Jussim, 1991; Willingham, Lewis, Morgan, & Ramist, 1990).

Not only do grades convey different types of information about the student but they also comprise different styles of grading across instructors (Gordon & Fay, 2010). For example, some instructors may choose to grade on a curve, which can take many forms but most commonly may involve the addition of the number of points between the highest exam score and 100 to all student exam scores in the class, or may entail transforming student grades or exam scores based on their relative position to the performance of other students in the class onto a normal distribution (Kulick & Wright, 2008). Other instructors may allow students to drop their lowest exam score from their final course grade or receive extra credit by participating in research studies (Gordon & Fay, 2010; Norcross, Dooley, & Stevenson, 1993).

Research has also found that instructor background can influence grading practices (Edwards, 2000; Hu, 2005; Stumpf & Freedman, 1979). Adjunct or part-time instructors tend to award higher grades to students than full-time faculty, which is likely related to their desire to obtain high student ratings to ensure that their contracts are renewed each year (Sonner, 2000). Greenwald and Gillmore (1997) noted that while giving students high grades is not sufficient to ensure high ratings, if an instructor varied nothing but student grading practices between two of the same course offerings, the students in the more leniently graded course would be expected to produce higher instructor ratings.

The academic discipline of the course can also impact the related grading practices. Many studies have found that grading practices tend to be stricter in science- and mathematics-related fields, while grading practices in the humanities and social sciences tend to be more lenient (Elliott & Strenta, 1988; Hu, 2005; Ramist et al., 1990; Rojstaczer & Healy, 2010; Sabot & Wakeman-Linn, 1991; Shaw & Patterson, 2010). Achen and Courant (2009) explain such differences by acknowledging that while students can debate the more subjective grades assigned to papers in an English composition course, “Students cannot easily quarrel with a determination that they failed to differentiate an exponential function or to reproduce a chemical formula on the midterm” (p. 87). In addition, to the extent that adjunct instructors tend to give higher grades to students than full-time faculty, the first-year writing courses in the English department are also more frequently taught by adjunct instructors (Avakian, 1995). These grading disparities not only threaten the validity of the FYGPA as a measure of academic performance in college but can cause serious consequences related to academic behaviors in higher education whereby students will actively seek out more leniently-graded major fields and/or instructors with lower grading standards (Geisinger, 1979; Hu, 2005).

Also of note, there have been some differences found in grading practices by institution type (Hu, 2005; Kuh & Hu, 1999). For example, private schools tend to award slightly higher grades than public institutions (Hu, 2005; Rojstaczer & Healy, 2010). Southern institutions tend to give lower grades than institutions in other regions of the U.S. (Rojstaczer & Healy, 2012). There are other studies that find that two-year institutions tend to award higher grades than four-year institutions (Friedl, Pittenger, & Sherman, 2012) while some show that first-year grades in community colleges tend to be lower than in four-year colleges (Adelman, 2004).

Trends in College Grading Practices

The majority of the research examining trends in college grading practices has been undertaken to study whether there is evidence of grade inflation, or whether there is an increase in grades without a corresponding increase in student ability (Bejar & Blew, 1981). However, Hu (2005) outlines major grading problems, including grade inflation, that are negatively impacting our understanding and the utility of college grades. First Hu discusses the issue of grade increases whereby averages grades in a particular course increase over time and this can provide evidence of appropriately increasing performance by students over time or, if unaccompanied by other indicators of achievement, could provide evidence of grade inflation. Grade inflation is noted as problematic because it will unfairly advantage more recent generations of students and can undermine the motivational function of grades in student learning. Grade compression is another issue offered by Hu that refers to the limited variation of grades awarded in particular courses (e.g., all students receiving A's and B's) that prohibits the differentiation of student performance and the inherent meaning of the grade, thereby reducing its use in understanding what a student is capable of, which is particularly problematic for graduate school or employment decisions after college. Finally, Hu proposed that grading disparity can present complications when grades are awarded in a different manner across disciplines, which can impact student course choices in college and therefore lead to GPA and course grade inflation.

Recent research on college grading practices across the U.S. has identified (e.g., Kuh & Hu, 1999; Rojstaczer & Healy, 2010, 2012) but also questioned (e.g., Adelman, 2004) the existence of increasing college grades as well as grade inflation. A comprehensive and often-referenced research project by Rojstaczer and Healy (2010, 2012) has found that across approximately 135 postsecondary institutions, A's represented 43% of all letter grades given in 2008, which is an increase in 28 percentage points from 1960 and 12 percentage points since 1988. While in the 1940s through the mid-1960s, the most common letter grade assigned was a C with about 35% of student grades at a C, by 2008 the percentage of C grades dropped to about 15%. Rojstaczer and Healy add that based on the small increase in SAT scores over this same time period, it does not appear that the rise in college grades is accompanied by the same increase in student achievement. Alternatively, Adelman (2004) has found less clear patterns in the distribution of student grades across the 1972, 1982, and 1992 12th graders who attended college. Adelman found that approximately 27%, 26%, and 28% of student grades were an A in 1972, 1982, and 1992, respectively. With regard to the average student GPA for students earning more than 10 credits, this fluctuated from a 2.70 in 1972, to a 2.66 in 1982, to a 2.74 in 1992. While the overall grading patterns were nuanced and complex over time, Adelman did find clear variation in the distribution of grades from course to course when he examined the largest volume courses for the 1992 cohort. For example, while 73% of Technical Writing grades were an A or B, only 45% of Calculus grades and U.S. Government grades were an A or B. Hu (2005) eloquently summarized much of the research on trends in college grades by acknowledging that while many college campuses have observed upward trends in student grades, the national research has shown smaller increases and changes. What does appear to more significantly threaten the meaning and utility of grades are the disparities found by course area that impact the courses that students choose to take and pursue for their studies and further provide incentives for faculty to lower their grading standards.

The Current Study

This study will explore recent trends in college grading practices, in general, and by academic discipline and institutional characteristics across a large and diverse sample of four-year institutions in the U.S. Results contextualize and improve our understanding and use of college grades—particularly as they are employed as the criteria in numerous studies on college readiness, admission test validity, and in other work related to educational benchmarking and accountability.

Method

Sample

The data from this study are from a longitudinal database developed to examine the validity of the SAT in partnership with four-year colleges and universities in the U.S. The students in this study entered college for the first time in fall 2007, 2008, 2009, 2010, or 2011. To be included in the study, students needed to have SAT scores and a first-year grade point average on record. This resulted in 638,197 students from 72 four-year institutions in the sample. The sample included more female (53.2%) than male students and more white students (67.2%) than those from other racial/ethnic groups. Students varied by academic ability, as shown by SAT scores, with the majority of students (40.1%) falling within the middle SAT score band of 1500–1790 (Table 1). Table 2 includes the characteristics of the 72 four-year institutions in this sample. There are slightly more private institutions (52.8%) than public, more moderately selective at 50%–75% admitted (52.8%) than under 50% admitted or over 75% admitted, and the most frequent institution size category is medium to large: 2,000 to 7,499 undergraduates (36.1%).

Student course-taking information and grades were provided by each institution. For this study, non-remedial courses taken by the student in the first year in college were examined. Non-remedial courses were grouped together by associated subject domain and a domain-specific grade point average was calculated. The domains of interest are business and communications, computer sciences, engineering, English, foreign and classical languages, history, humanities, mathematics, natural sciences, health sciences, and social sciences.

Measures

SAT Scores. A student’s most recent SAT scores from administrations prior to the redesign of the SAT were obtained for each student in the study spanning five cohorts.

SAT Questionnaire Responses. Self-reported gender, race/ethnicity, best language, and highest parental education level were obtained from the SAT Questionnaire that students complete during registration for the SAT.

First-Year GPA. Each participating institution supplied first-year grade point average (FYGPA) values for students included in this sample.

College Grades. First-year GPAs and grades in all courses in the first year of college were obtained from each student's higher education institution. All courses were coded for content area so that analyses could be conducted on course-specific grade point averages. Domain-specific grade point averages were calculated within student, across all relevant course grades received in a particular area during the first year of college (excluding remedial coursework). For example, if a student took only one mathematics course in his or her first year, then his or her average course grade in mathematics is based on the grade earned in that one course. If the student took three mathematics courses, the average course grade is based on the average of the three course grades earned (taking into account the grades earned and the number of credits associated with each grade).

Analyses

The focus of the current study is to investigate the differences in college grading practices by student and institutional characteristics, as well as by academic discipline, as course grades and first-year grade point average (FYGPA) tend to be the most commonly utilized criteria in admission validity studies and college readiness research. In addition, trends in college grades were examined over a five-year period to determine the stability of these grading differences. Domain-specific grade point averages were calculated using non-remedial, first-year grades for each student. The domains were business and communications, computer sciences, engineering, English, foreign and classical languages, history, humanities, mathematics, natural sciences, health sciences, and social sciences.

Descriptive analyses were used to compare means and standard deviations of FYGPAs and domain-specific GPAs across five first-year cohorts. Mean domain-specific grades were compared, controlling for students' SAT scores as well as institutional selectivity. Controlling for students' SAT scores allows for comparisons of domain-specific GPA that are independent of student ability. To remove institutional effects as a contributor to grading differences, institutional selectivity was included as an additional control in analyzing mean grades by discipline and SAT score band. Finally, at the institution level, a student's FYGPA was predicted using their SAT scores and high school GPA. This predicted FYGPA was then compared to the average course grade in each domain to understand most generally how the inclusion of grades in different disciplines in a FYGPA will tend to impact differential validity and prediction analyses.

Results

Table 3 shows that there are slight increases in overall FYGPA and by subgroup over time. The average FYGPA for the 2007 entering cohort was 2.95 (SD = 0.76) and in 2011, the average FYGPA for the entering cohort was 3.01 (SD = 0.72). Females tended to have higher FYGPAs than males, but there were increases in FYGPA for both females and males from 2007 to 2011. Asian, Asian American, or Pacific Islander students tended to have the highest FYGPAs (in 2011, M = 3.13, SD = 0.64) across the racial/ethnic subgroups, followed by white students (in 2011, M=3.08; SD=0.69), Other students (in 2011, M = 2.99, SD = 0.73), Hispanic, Latino, or Latin American students (in 2011, M = 2.80, SD = 0.77), American Indian or Alaska Native students (in 2011, M = 2.76, SD = 0.78), and black or African American students (in 2011, M = 2.56, SD = 0.81). With regard to best language, while the vast majority of students were in the English Only language group, the students in the smaller Another Language group tended to have the highest FYGPAs. They also experienced the largest increases in FYGPA across the three language groups from 2.99 (SD = 0.73) in 2007 to 3.10 (SD = 0.68) in 2011. As one would intuitively expect, there were differences in FYGPA by SAT score band whereby higher SAT score bands have higher mean FYGPAs and the differences are quite pronounced. For example, in 2011, students in the 600–1190 score band have a mean FYGPA of 2.28 (SD = 0.81) while students in the 2100–2400 score band have a mean FYGPA of 3.51 (SD = 0.47). The FYGPAs in the lower score bands increased slightly more than those in the higher score bands over the five years of this study. With regard to highest parental education level, those students whose parents obtained less than a bachelor's degree tended to have the lowest mean FYGPAs (in 2011, M = 2.79, SD = 0.80) while those whose parents obtained more than a bachelor's degree had the highest (in 2011, M = 3.17, SD = 0.63). The three parental education level subgroups had similar increases in FYGPA from 2007 to 2011.

Similar to the FYGPA variation evident by student subgroups, mean FYGPAs also varied by institutional subgroup and remained relatively consistent across cohort years (see Table 4). For example, mean FYGPA tends to be higher at private (in 2011, M = 3.17, SD = 0.57) versus public institutions (in 2011, M = 2.95; SD = 0.76) and higher at the most selective institutions (in 2011, M = 3.20; SD = 0.54) than the least selective institutions (in 2011, M = 2.80; SD = 0.77). With regard to institution size, there were few clear patterns in FYGPA differences, though small institutions seemed to have the largest increase in mean FYGPA from 2007 to 2011 (in 2007, M = 2.87, SD = 0.71; in 2011, M = 2.97, SD = 0.66) however they also had the lowest mean FYGPA to start.

Table 5 displays the average domain-specific GPA by discipline and the percentage of students taking at least one course in each domain over time. The most popular subject domains across all years were social sciences (ranging from 81.1% to 81.9% of students taking at least one course), mathematics (69.9%–70.6%), English (68.3%–71.1%), and natural sciences (66.9%–68.3%). The least popular subject domain was engineering with only 11.6% to 12.6% of students taking at least one course in that area. As further shown in

Figure 1 (based on the values in Table 5) there are major differences in average course grades by discipline. Disciplines such as mathematics (in 2011, M = 2.70, SD = 1.06) and natural sciences (in 2011, M = 2.75, SD = 0.97) have lower mean grades and larger standard deviations than disciplines like foreign and classical languages (in 2011; M = 3.23, SD = 0.83) and health sciences (in 2011; M = 3.19, SD = 0.90). Most discipline-specific GPAs remained relatively stable from 2007 to 2011, though the lower GPAs in 2007 (e.g., mathematics, natural sciences, history, social sciences) tended to have slightly larger increases in GPA to 2011.

The next natural investigation is to understand whether there are differences in ability level (rather than just in grading practices) within domain. Figure 2 shows the relationship between average SAT score and average course grade within a domain across the entire sample (all five cohorts together). From this figure, it can be seen that certain domains have higher average SAT scores and higher course grade averages (foreign and classical languages and engineering). Yet other domains such as mathematics and natural sciences have relatively high SAT scores, yet the lowest course grade averages. Further, domains such as English and health sciences have high course grades and moderate SAT scores.

In addition to observing grading differences and SAT score differences by domain, Figure 3 shows that grading differences by domain remain evident even when controlling for prior academic ability measured by SAT score band. While grades within discipline are higher as SAT score bands increase, we also see that students within the same SAT score band across disciplines are earning different average domain-specific GPAs, dependent on the discipline. For example, in the middle score band (1500–1790) students in mathematics earn a 2.59 (SD = 1.07) course GPA, on average, whereas students in English earn a 3.17 (SD = 0.83) course GPA, on average (see Table 6). Patterns like this are seen in all score bands across the domain-specific GPAs. This figure also shows that there tends to be greater differentiation in average course grade by SAT score band in domains such as history, humanities, or natural sciences; however, average course grades appear more compressed by SAT score band within engineering.

When institutional selectivity is considered along with SAT score band and domain, we see an institutional effect whereby higher grades are generally more prevalent at selective institutions within the same domain and SAT score band (see Table 7). As selectivity decreases, mean grades also decrease. For example, in the 1500–1790 SAT score band within mathematics, students had an average course GPA of 2.65 (SD = 0.95) in the under 50% admitted group, 2.60 (SD = 1.06) in the 50%–75% admitted group, and 2.51 (SD = 1.17) in the over 75% admitted group. However, these patterns can switch based on score band. For example, students within the 1200–1490 score band in the English domain have average GPAs of 2.96 (SD = 0.74), 2.80 (SD = 0.98), and 2.81 (SD = 0.96), as selectivity decreases. Yet, as selectivity decreases for those in the highest score band (2100–2400), English GPAs seem to increase with students earning a 3.45 (SD = 0.65), 3.55 (SD = 0.70), and 3.63 (SD = 0.66) average GPA, respectively. This was also the trend for students in the highest score SAT band in engineering, mathematics, and natural sciences, among other fields.

In addition to looking at FYGPA and domain-specific GPAs over time and by various subgroups, it is useful to consider the amount of over- and under-prediction by domain as related to expected college performance. This addresses some of the thorny prediction problems due to ‘‘noise’’ in the criterion based on differences in course-taking in college. Table 8 shows the average difference between the domain-specific GPA and the students’ predicted FYGPAs to understand how different (and in what direction) the average course grade is from how a student was generally expected to perform at an institution. Predicted FYGPA was used because the interest is in generally understanding how a student was expected to perform at the institution versus how the domain of interest assigns grades, on average. Predicted FYGPA was determined within institution from a student’s HSGPA and all three sections of the SAT. Positive differences indicate that on average, course grades were higher than the predicted FYGPA and negative differences indicate the reverse. Four of the domains had negative differences. Social sciences and history both had small differences (-0.07 and -0.10, respectively) whereas the difference was much greater in natural sciences (-0.29) and mathematics (-0.37). This suggests that natural sciences and mathematics grade more harshly than the others, especially given that the same student could be captured in multiple subject areas and have the same predicted FYGPA regardless of which area was being examined. Health sciences (0.23), English (0.18), and business and communications (0.15) had course grades that were higher than predicted for students, indicating that there are domain-specific grading tendencies that are likely not related to actual performance but rather to the academic/departmental grading culture and these differences will impact predictive models of college success when FYGPA is used as the criterion.

Discussion

Kostal, Kuncel, and Sackett (2016) note that college grades serve a number of important purposes that necessitate their periodic analysis. For example, grades allow students to understand how well they have mastered material and the extent to which they may need to apply additional effort to increase mastery. Grades can influence future course and career choices by signaling a level of success in a particular domain, and they can also inform graduate school admission and hiring decisions by serving as an endorsement of the level of knowledge or success a student has achieved. Yet, we also know that grades can include information other than subject mastery, such as student level of effort, circumstances, conscientiousness, or instructor attitudes or biases (Brookhart, 1993; Harris, 1940; Jussim, 1991; Willingham, Lewis, Morgan, & Ramist, 1990). The current study undertook an exploratory, longitudinal analysis of college grades to better understand recent trends, the role of student and institutional characteristics, as well as the role of academic discipline. Findings from this study of 72 four-year institutions show that there were small increases in overall FYGPA from 2007 to 2011 and that FYGPA tended to vary by student and institutional subgroups. There were also major differences in average course grade by academic discipline and these differences essentially held after controlling for students' SAT scores and institutional selectivity. These findings have implications for the general interpretation of college grades and the more specific use of college grades as criteria in validity research and educational benchmarking work.

The overall FYGPA in this study increased from 2.95 in 2007 to 3.01 in 2011, a period of five years. These FYGPA increases were evident in most student (demographic and academic) and institutional subgroups and seem to be in keeping with Kostal et al.'s (2016) .08 estimate of the size of grade increase across a decade. The FYGPA increases in this study can be further contextualized by stagnant or slightly decreasing average SAT scores during the same period, which were 501, 514, and 493 for Critical Reading, Mathematics, and Writing, respectively, in 2007, and were 497, 514, and 489 for Critical Reading, Mathematics, and Writing, respectively, in 2011 (College Board, 2011).

The FYGPA patterns by student subgroup mirror subgroup differences found in most academic measures (Camara & Schmidt, 1999; Kobrin, Sathy, & Shaw, 2006). Though it is possible that these FYGPA subgroup differences are not receiving the attention they should. For example, there are frequent claims in the press of racial bias on standardized tests (e.g., Jaschik, 2010); however, the existence of subgroup differences on an educational measure does not signal bias in the measure in and of itself. It does, however, signal something to be mindful of and monitor. With regard to FYGPA, in 2011, we see black or African American students earning an average FYGPA of 2.56. White students in that same year earned an average FYGPA of 3.08, representing a difference of .52 on the GPA scale. These lower mean GPAs will likely have related implications for retention, completion, and future job and graduate school prospects (Kopp & Shaw, 2016; Kostal et al., 2016). These lower average FYGPAs also have implications for differential validity and prediction research. For example, findings over time have consistently shown that test scores and high school grades overpredict the performance of black students in college (e.g., Mattern & Patterson, 2014). Instead of solely looking for issues with the predictors (or tests) in such scenarios, it would seem useful to have institutions look inward to understand why certain groups of students are exhibiting much lower performance in college than expected and performing much more poorly than other groups of students and determine whether certain supports can be put in place to improve college success.

Some of the institutional differences in FYGPA were also quite pronounced in that, on average, students can expect to receive higher FYGPAs at private institutions and selective institutions. However, when controlling for selectivity, SAT score band, and academic domain, more nuanced patterns were seen. In some academic departments, there were average course grade increases or decreases with increasing selectivity that could change dependent on SAT score band. An example was seen in the average mathematics GPA, which increased with decreasing selectivity (from most to least selective) for the highest SAT score band (2100–2400) from 3.10 to 3.26, while the mathematics GPA decreased with decreasing selectivity for the students in the 1200–1490 SAT score band from 2.31 to 2.20.

As expected, there were major differences in average course grade by academic discipline. For example, the average mathematics GPA in 2011 was 2.70 while the average English GPA in 2011 was 3.14. These differences held when examined by SAT score band. For students in the 2100–2400 SAT score band, the mathematics GPA was 3.13 and the English GPA was 3.49. Within academic domain there were also differences in the amount of average grade increase evident from 2007 to 2011, whereby slightly greater average grade increases were seen in the more harshly graded fields of mathematics, natural sciences, and history. There were also differences in the spread of average grades earned by SAT score band.

All of these grading differences will impact research that relies on FYGPA as the outcome measure (Keiser, Sackett, Kuncel, & Brothen, 2016; Mattern, Sanchez, & Ndum, 2017). The study analysis examining differences between predicted student performance and actual performance by domain confirms this by showing that in certain academic domains, students are earning much higher (e.g., English, business or communications) or much lower (e.g., mathematics, natural sciences) grades than would be expected based on their previous performance. One implication of this finding is that with regard to admission validity research, it may be worthwhile to consider students' intended major in the predictive model or have different models depending on major, as expected performance will vary just by the nature of the domain of coursework pursued.

These grading differences by discipline can also explain why college readiness benchmarks that take into account the academic domain of interest will result, relatively, in scores representing those benchmarks. For example, the SAT College and Career Readiness benchmark (related to a 75% probability of earning at least a C in first-semester, credit-bearing college courses in the domain) in Evidence-Based Reading and Writing is 480 while the benchmark for Math is 530. This pertains to the findings from this study showing that, although average English course grades increase with SAT score band, even students in the lower SAT score bands will tend to receive strong English grades in college. The same cannot be said for mathematics or for a number of other disciplines. However, the key takeaway is that this will vary by discipline.

There are a few limitations of this study that are worth noting. The sample includes those students at institutions who took the SAT and therefore does not include every student within an entering cohort. Because of the large number of institutions and students in this sample, this should not present a major interpretational problem, but it is worth noting. Also, it would have been useful to include two-year institutions in this study to further analyze grading differences by two- versus four-year institutions and understand how grades within an academic domain might vary by institution in this way.

Future research in this area should consider follow-up surveys by academic departments to better understand their grading practices, considerations, and communications. Also, future research should examine how these results may change when the outcome or criteria is cumulative GPA. It would also be useful to compare the relationship between expected student performance based on HSGPA alone and expected student performance based on SAT alone with the actual performance in each domain to determine whether the SAT or HSGPA are more or less accurate for predicting performance in different disciplines. Research may also want to delve into the role of institutional selectivity in grading practices by discipline, as this study found some unusual findings in that area. Finally, we think it is worthwhile for additional research to better understand the racial/ethnic differences observed in FYGPA, and in particular the lower FYGPAs for black or African American students. It would be useful to consider how improved preparation or supports on campus may result in stronger college performance and ultimately more long-term positive educational outcomes.

References

Achen, A. C., & Courant, P. N. (2009). What are grades made of? *Journal of Economic Perspectives, 23*(3), 77–92.

Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972–2000*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218–223.

Avakian, A. N. (1995). Conflicting demands for adjunct faculty. *Community College Journal, 65*(6), 34–36.

Bejar, I. I., & Blew, E. O. (1981). Grade inflation and the validity of the Scholastic Aptitude Test. *American Educational Research Journal, 18*(2), 143-156.

Berry, C. M., & Sackett, P. R. (2009). Individual differences in course choice result in underestimation of college admissions system validity. *Psychological Science, 20*, 822–830.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123–142.

Camara, W. J. (2005). Broadening predictors of college success. In W. J. Camara and E. W. Kimmel (Eds.), *Choosing students: Higher education admissions tools for the 21st century* (pp. 81–105). Mahwah, NJ: Lawrence Erlbaum Associates.

Camara, W. J., & Echternacht, G. (2000). *The SAT I and High School Grades: Utility in Predicting Success in College* (College Board Research Note RN-10). New York: The College Board.

Camara, W. J., & Quenemoen, R. (2012). *Defining and measuring college and career readiness and informing the development of performance level descriptors (PLDs).* Commissioned white paper for PARCC. Available at http://www.parcconline.org/sites/parcc/files/PARCC%20CCR%20paper%20v14%201-8-12.pdf

Camara, W. J., & Schmidt, A. E. (1999). *Group Differences in Standardized Testing and Social Stratification* (College Board Research Report No. 99-5). New York: The College Board.

Edwards, C. H. (2000). Grade inflation: The effects on educational quality and personal well-being. *Education, 120*(3), 538.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement, 25*(4), 333–347.

Friedl, J., Pittenger, D. J., & Sherman, M. (2012). Grading standards and student performance in community college and university courses. *College Student Journal, 46*(3), 526–532.

Geisinger, K. F. (1979). A note on grading policies and grade inflation. *Improving College and University Teaching, 27*, 113–115.

Gordon, M. E., & Fay, C. H. (2010). The effects of grading and teaching practices on students' perceptions of grading fairness. *College Teaching, 58*(3), 93–98.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*, 1209–1217.

Harris, D. (1940). Factors affecting college grades: A review of the literature, 1930–1937. *Psychological Bulletin, 37*(3), 125–165.

Hu, S. (2005). *Beyond grade inflation: Grading problems in higher education.* San Francisco, CA: Jossey-Bass.

Jaschik, S. (2010, June 21). New evidence of racial bias on SAT. *Inside Higher Ed.* Retrieved from http://www.insidehighered.com

Jussim, L. (1991). Grades may reflect more than performance: Comment on Wentzel (1989). *Journal of Educational Psychology, 83*(1), 153–155.

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology, 101*(4), 569–581.

Kobrin, J. L., Sathy, V., & Shaw, E. J. (2006*). A Historical View of Subgroup Performance Differences on the SAT* (College Board Research Report 2006-5). New York: The College Board.

Kopp, J. P., & Shaw, E. J. (2016). How final is leaving college while in academic jeopardy? Examining the utility of differentiating college leavers by academic standing. *Journal of College Student Retention: Research, Theory & Practice 18*(1), 2–30.

Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice, 35*(1), 11–20.

Kuh, G. D., & Hu, S. (1999). Unraveling the complexity of the increase in college grades from the mid-1980s to the mid-1990s. *Educational Evaluation & Policy Analysis, 21*(3), 297–320.

Kulick, G., & Wright, R. (2008). The Impact of Grading on the Curve: A Simulation Analysis. *International Journal for the Scholarship of Teaching and Learning, 2*(2), 1-25.

Mattern, K. D., & Patterson, B. F. (2014). Synthesis of recent SAT validity findings: Trend Data over time and cohorts. New York: The College Board. Retrieved from http://files.eric.ed.gov/fulltext/ED556462.pdf

Mattern, K. D., Sanchez, E., & Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice 0*(0), 1–11.

Norcross, J. C., Dooley, H. S., & Stevenson, J. F. (1993). Faculty use and justification of extra credit: No middle ground? *Teaching of Psychology, 20*(4), 240–242

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT. In W. W. Willingham. C. Lewis, R. Morgan, & L. Ramist (Eds.), *Predicting college grades: An analysis of institutional trends over two decades* (pp. 253–288). Princeton, NJ: Educational Testing Service.

Rojstaczer, S., & Healy, C. (2010). Grading in American colleges and universities. *Teachers College Record*, ID Number: 15928.

Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940–2009. *Teachers College Record, 114*(7), 1–23.

Sabot, R., & Wakeman-Linn, J. (1991). Grade inflation and course choice. *Journal of Economic Perspectives, 5*(1), 159–70.

Shaw, E. J., & Patterson, B. F. (2010). *What Should Students Be Ready For in College? A Look at First-Year Course Work in Four-Year Postsecondary Institutions in the U.S.* (College Board Research Report 2010-1). New York: The College Board.

Sonner, B. S. (2000). A Is for "Adjunct": Examining grade Inflation in higher education. *Journal of Education for Business, 76*(1), 5–9.

Stumpf, S. A., & Freedman, R. D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology,* 71, 293–302.

The College Board. (2011). *2011 College-bound seniors: Total group profile report. New York: The College Board*. Retrieved from: http://media.collegeboard.com/digitalServices/pdf/research/cbs2011_total_group_report.pdf

Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Educational Testing Service.

Table 1: Student Demographic Characteristics from 2007–2011 and Overall

		2007		2008		2009		2010		2011		All Cohorts	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender	Male	58,695	47.0	59,458	46.9	59,829	47.2	58,939	46.7	61,670	46.2	298,591	46.8
	Female	66,105	53.0	67,338	53.1	67,046	52.8	67,388	53.3	71,729	53.8	339,606	53.2
Race/Ethnicity	American Indian or Alaska Native	664	0.5	603	0.5	602	0.5	535	0.4	557	0.4	2,961	0.5
	Asian, Asian American, or Pacific Islander	10,940	8.8	12,653	10.0	13,214	10.4	13,869	11.0	15,558	11.7	66,234	10.4
	Black or African American	7,037	5.6	7,521	5.9	7,862	6.2	8,472	6.7	9,118	6.8	40,010	6.3
	Hispanic, Latino, or Latin American	10,560	8.5	11,834	9.3	12,867	10.1	13,690	10.8	14,946	11.2	63,897	10.0
	No Response	7,355	5.9	4,009	3.2	3,236	2.6	3,594	2.8	2,048	1.5	20,242	3.2
	Other	3,379	2.7	3,231	2.5	3,187	2.5	2,801	2.2	3,481	2.6	16,079	2.5
	White	84,865	68.0	86,945	68.6	85,907	67.7	83,366	66.0	87,691	65.7	428,774	67.2
Best Language	English Only	113,498	90.9	105,390	83.1	112,822	88.9	107,540	85.1	112,206	84.1	551,456	86.4
	English and Another Language	6,602	5.3	7,695	6.1	9,149	7.2	14,619	11.6	17,113	12.8	55,178	8.6
	Another Language	1,227	1.0	1,836	1.4	2,191	1.7	2,452	1.9	2,706	2.0	10,412	1.6
	Not Stated	3,473	2.8	11,875	9.4	2,713	2.1	1,716	1.4	1,374	1.0	21,151	3.3
SAT Score Band	SAT Score Band 1 (600–1190)	3,653	2.9	3,819	3.0	3,919	3.1	3,992	3.2	4,167	3.1	19,550	3.1
	SAT Score Band 2 (1200–1490)	25,204	20.2	25,127	19.8	24,599	19.4	24,636	19.5	27,018	20.3	126,584	19.8
	SAT Score Band 3 (1500–1790)	50,584	40.5	51,092	40.3	50,595	39.9	50,376	39.9	52,989	39.7	255,636	40.1
	SAT Score Band 4 (1800–2090)	37,978	30.4	39,123	30.9	39,787	31.4	38,968	30.8	40,358	30.3	196,214	30.7
	SAT Score Band 5 (2100–2400)	7,381	5.9	7,635	6.0	7,975	6.3	8,355	6.6	8,867	6.6	40,213	6.3
Highest Parental Education	< Bachelor's Degree	35,018	28.1	33,923	26.8	35,570	28.0	35,232	27.9	38,526	28.9	178,269	27.9
	Bachelor's Degree	40,636	32.6	38,456	30.3	40,706	32.1	41,861	33.1	46,948	35.2	208,607	32.7
	> Bachelor's Degree	41,230	33.0	38,062	30.0	40,650	32.0	42,023	33.3	43,009	32.2	204,974	32.1
	No Response	7,916	6.3	16,355	12.9	9,949	7.8	7,211	5.7	4,916	3.7	46,347	7.3
Overall		124,800	100.0	126,796	100.0	126,875	100.0	126,327	100.0	133,399	100.0	638,197	100.0

Table 2: Sample Institutional Characteristics

		k	% Sample
Control	Public	34	47.2
	Private	38	52.8
Selectivity	Under 50% Admitted	20	27.8
	50%–75% Admitted	38	52.8
	Over 75% Admitted	14	19.4
Size	Small	12	16.7
	Medium to Large	26	36.1
	Large	14	19.4
	Very Large	20	27.8
Overall		72	100.0

Note: *k* represents the number of institutions. With regard to institution size, small = 750 to 1,999 undergraduates; medium to large = 2,000 to 7,499 undergraduates; large = 7,500 to 14,999 undergraduates; and very large = 15,000 or more undergraduates.

Table 3: Average FYGPA from 2007–2011, by Student Characteristics

		2007		2008		2009		2010		2011	
		n	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**
Gender	Male	58,695	2.86 (0.80)	59,458	2.89 (0.79)	59,829	2.88 (0.79)	58,939	2.92 (0.74)	61,670	2.93 (0.75)
	Female	66,105	3.03 (0.72)	67,338	3.05 (0.71)	67,046	3.04 (0.72)	67,388	3.07 (0.68)	71,729	3.08 (0.68)
Race/Ethnicity	American Indian or Alaska Native	664	2.78 (0.81)	603	2.77 (0.84)	602	2.73 (0.83)	535	2.84 (0.81)	557	2.76 (0.78)
	Asian, Asian American, or Pacific Islander	10,940	3.05 (0.70)	12,653	3.08 (0.68)	13,214	3.06 (0.69)	13,869	3.11 (0.64)	15,558	3.13 (0.64)
	Black or African American	7,037	2.51 (0.82)	7,521	2.51 (0.82)	7,862	2.50 (0.84)	8,472	2.55 (0.79)	9,118	2.56 (0.81)
	Hispanic, Latino, or Latin American	10,560	2.69 (0.81)	11,834	2.73 (0.80)	12,867	2.71 (0.80)	13,690	2.77 (0.78)	14,946	2.80 (0.77)
	No Response	7,355	2.93 (0.78)	4,009	2.95 (0.79)	3,236	2.97 (0.78)	3,594	3.03 (0.72)	2,048	2.94 (0.75)
	Other	3,379	2.96 (0.73)	3,231	2.99 (0.73)	3,187	2.96 (0.76)	2,801	2.98 (0.73)	3,481	2.99 (0.73)
	White	84,865	3.00 (0.74)	86,945	3.03 (0.73)	85,907	3.03 (0.73)	83,366	3.07 (0.68)	87,691	3.08 (0.69)
Best Language	English Only	113,498	2.96 (0.76)	105,390	2.97 (0.75)	112,822	2.97 (0.75)	107,540	3.02 (0.71)	112,206	3.02 (0.72)
	English and Another Language	6,602	2.81 (0.79)	7,695	2.83 (0.78)	9,149	2.82 (0.78)	14,619	2.87 (0.75)	17,113	2.91 (0.74)
	Another Language	1,227	2.99 (0.73)	1,836	3.02 (0.73)	2,191	3.02 (0.73)	2,452	3.08 (0.68)	2,706	3.10 (0.68)
	Not Stated	3,473	2.80 (0.83)	11,875	3.12 (0.69)	2,713	3.01 (0.77)	1,716	2.97 (0.74)	1,374	2.98 (0.73)
SAT Score Band	SAT Score Band 1 (600–1190)	3,653	2.19 (0.85)	3,819	2.23 (0.85)	3,919	2.19 (0.83)	3,992	2.25 (0.81)	4,167	2.28 (0.81)
	SAT Score Band 2 (1200–1490)	25,204	2.54 (0.79)	25,127	2.58 (0.79)	24,599	2.56 (0.80)	24,636	2.60 (0.77)	27,018	2.61 (0.78)
	SAT Score Band 3 (1500–1790)	50,584	2.91 (0.72)	51,092	2.93 (0.71)	50,595	2.92 (0.71)	50,376	2.98 (0.66)	52,989	2.98 (0.67)
	SAT Score Band 4 (1800–2090)	37,978	3.23 (0.62)	39,123	3.25 (0.60)	39,787	3.23 (0.62)	38,968	3.26 (0.56)	40,358	3.28 (0.56)
	SAT Score Band 5 (2100–2400)	7,381	3.49 (0.51)	7,635	3.50 (0.51)	7,975	3.49 (0.52)	8,355	3.51 (0.46)	8,867	3.51 (0.47)

Highest Parental Education	< Bachelor's Degree	35,018	2.71 (0.84)	33,923	2.73 (0.83)	35,570	2.73 (0.83)	35,232	2.77 (0.80)	38,526	2.79 (0.80)
	Bachelor's Degree	40,636	2.99 (0.72)	38,456	3.00 (0.72)	40,706	3.01 (0.72)	41,861	3.05 (0.67)	46,948	3.06 (0.68)
	> Bachelor's Degree	41,230	3.11 (0.67)	38,062	3.12 (0.67)	40,650	3.12 (0.67)	42,023	3.17 (0.62)	43,009	3.17 (0.63)
	No Response	7,916	2.89 (0.78)	16,355	3.06 (0.72)	9,949	2.96 (0.77)	7,211	2.96 (0.73)	4,916	2.87 (0.75)
Overall		124,800	2.95 (0.76)	126,796	2.97 (0.75)	126,875	2.96 (0.76)	126,327	3.00 (0.71)	133,399	3.01 (0.72)

Table 4: Average FYGPA from 2007–2011, by Institutional Characteristics

		2007		2008		2009		2010		2011	
		n	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**
Control	Public	90,908	2.89 (0.80)	92,034	2.92 (0.79)	92,314	2.91 (0.79)	91,318	2.95 (0.75)	98,307	2.95 (0.76)
	Private	33,892	3.11 (0.62)	34,762	3.12 (0.62)	34,561	3.11 (0.63)	35,009	3.15 (0.58)	35,092	3.17 (0.57)
Selectivity	Under 50% Admitted	14,582	3.23 (0.53)	31,262	3.20 (0.57)	26,953	3.17 (0.60)	22,517	3.21 (0.54)	36,077	3.20 (0.54)
	50%–75% Admitted	83,610	2.96 (0.75)	77,881	2.93 (0.77)	77,087	2.92 (0.78)	90,480	2.98 (0.73)	79,544	2.97 (0.76)
	Over 75% Admitted	26,608	2.74 (0.83)	17,653	2.76 (0.86)	22,835	2.85 (0.79)	13,330	2.81 (0.77)	17,778	2.80 (0.77)
Size	Small	3,852	2.87 (0.71)	3,941	2.92 (0.72)	3,800	2.87 (0.76)	3,944	2.96 (0.69)	3,881	2.97 (0.66)
	Medium to Large	21,539	2.98 (0.77)	19,157	3.02 (0.76)	16,399	3.03 (0.76)	16,387	3.06 (0.73)	18,456	3.04 (0.76)
	Large	24,913	2.92 (0.77)	24,887	2.94 (0.76)	24,762	2.93 (0.76)	24,999	2.95 (0.74)	31,239	3.01 (0.69)
	Very Large	74,496	2.95 (0.76)	78,811	2.97 (0.75)	81,914	2.96 (0.75)	80,997	3.01 (0.70)	79,823	3.00 (0.72)
Overall		124,800	2.95 (0.76)	126,796	2.97 (0.75)	126,875	2.96 (0.76)	126,327	3.00 (0.71)	133,399	3.01 (0.72)

Note: With regard to institution size, small = 750 to 1,999 undergraduates; medium to large = 2,000 to 7,499 undergraduates; large = 7,500 to 14,999 undergraduates; and very large = 15,000 or more undergraduates.

Table 5: Average Course Grade by Domain-Specific Discipline

	2007			2008			2009			2010			2011		
	n	**% of total**	**M (SD)**	***n***	**% of total**	**M (SD)**	***n***	**% of total**	**M (SD)**	***n***	**% of total**	**M (SD)**	***n***	**% of total**	**M (SD)**
Business and Communications	42,327	33.9	3.05 (0.90)	44,815	35.3	3.07 (0.89)	42,325	33.4	3.07 (0.89)	43,584	34.5	3.10 (0.87)	47,308	35.5	3.09 (0.89)
Computer Sciences	17,661	14.2	3.00 (1.10)	17,807	14.0	3.01 (1.08)	17,442	13.7	3.03 (1.05)	17,443	13.8	3.10 (1.01)	19,113	14.3	3.06 (1.00)
Engineering	14,451	11.6	3.13 (0.92)	14,971	11.8	3.11 (0.91)	15,851	12.5	3.18 (0.87)	15,920	12.6	3.14 (0.86)	16,644	12.5	3.18 (0.83)
English	88,739	71.1	3.13 (0.87)	88,641	69.9	3.13 (0.88)	86,802	68.4	3.13 (0.88)	86,287	68.3	3.15 (0.86)	91,198	68.4	3.14 (0.88)
Foreign and Classical Languages	35,286	28.3	3.14 (0.89)	36,026	28.4	3.20 (0.87)	34,961	27.6	3.20 (0.86)	34,796	27.5	3.23 (0.83)	34,917	26.2	3.23 (0.83)
History	42,006	33.7	2.73 (1.03)	41,252	32.5	2.78 (1.02)	41,028	32.3	2.77 (1.01)	38,711	30.6	2.83 (1.00)	40,571	30.4	2.85 (1.01)
Humanities	41,144	33.0	3.05 (0.91)	41,389	32.6	3.06 (0.90)	41,190	32.5	3.06 (0.91)	40,875	32.4	3.06 (0.91)	42,432	31.8	3.07 (0.90)
Mathematics	87,574	70.2	2.56 (1.12)	89,007	70.2	2.62 (1.10)	88,680	69.9	2.64 (1.08)	88,551	70.1	2.68 (1.06)	94,221	70.6	2.70 (1.06)
Natural Sciences	83,466	66.9	2.69 (0.99)	85,185	67.2	2.71 (0.99)	86,491	68.2	2.72 (0.99)	86,211	68.2	2.75 (0.96)	91,046	68.3	2.75 (0.97)
Health Sciences	14,956	12.0	3.14 (0.94)	16,072	12.7	3.16 (0.93)	17,301	13.6	3.19 (0.91)	17,142	13.6	3.13 (0.92)	18,547	13.9	3.19 (0.90)
Social Sciences	101,193	81.1	2.85 (0.95)	103,394	81.5	2.89 (0.93)	103,475	81.6	2.90 (0.93)	103,524	81.9	2.94 (0.91)	108,579	81.4	2.94 (0.91)

Table 6: Average Course Grade by Domain-Specific Discipline, Controlling for Students' SAT Scores

	Business and Communications		Computer Sciences		Engineering		English		Foreign and Classical Languages		History	
	n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)
600–1190	8,307	2.46 (1.08)	2,568	2.28 (1.25)	587	2.66 (1.15)	15,472	2.39 (1.05)	2,533	2.41 (1.14)	7,760	1.93 (1.14)
1200–1490	52,565	2.75 (0.98)	16,072	2.72 (1.17)	6,876	2.81 (1.04)	96,904	2.81 (0.96)	22,663	2.76 (1.00)	52,962	2.38 (1.07)
1500–1790	93,400	3.09 (0.85)	33,890	3.04 (1.04)	26,940	3.04 (0.92)	180,160	3.17 (0.83)	64,052	3.11 (0.85)	85,437	2.83 (0.95)
1800–2090	56,804	3.37 (0.70)	29,913	3.19 (0.94)	34,181	3.24 (0.81)	125,825	3.37 (0.72)	69,610	3.37 (0.74)	48,818	3.19 (0.80)
2100–2400	9,283	3.55 (0.61)	7,023	3.44 (0.80)	9,253	3.41 (0.71)	23,306	3.49 (0.67)	17,128	3.55 (0.68)	8,591	3.44 (0.70)
Total Group	220,359	3.08 (0.89)	89,466	3.04 (1.05)	77,837	3.15 (0.88)	441,667	3.14 (0.87)	175,986	3.2 (0.86)	203,568	2.79 (1.01)

	Humanities		Mathematics		Natural Sciences		Health Sciences		Social Sciences	
	n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)
600–1190	4,529	2.18 (1.12)	9,637	1.93 (1.19)	9,296	1.77 (1.11)	3,290	2.53 (1.20)	16,280	2.08 (1.06)
1200–1490	34,278	2.60 (1.03)	81,452	2.24 (1.15)	76,791	2.24 (1.04)	21,655	2.92 (1.03)	106,507	2.47 (0.99)
1500–1790	82,752	2.99 (0.89)	186,618	2.59 (1.07)	177,362	2.68 (0.94)	36,081	3.19 (0.86)	211,343	2.89 (0.88)
1800–2090	69,656	3.31 (0.72)	142,321	2.89 (0.98)	140,600	3.00 (0.85)	19,935	3.41 (0.74)	155,454	3.21 (0.77)
2100–2400	15,815	3.55 (0.59)	28,005	3.13 (0.92)	28,350	3.24 (0.80)	3,057	3.61 (0.66)	30,581	3.43 (0.70)
Total Group	207,030	3.06 (0.90)	448,033	2.64 (1.08)	432,399	2.73 (0.98)	84,018	3.16 (0.92)	520,165	2.90 (0.93)

Table 7: Average Course Grade by Domain-Specific Discipline and Institution Selectivity, Controlling for Students' SAT Scores

		Business and Communications		Computer Sciences		Engineering		English		Foreign and Classical Languages		History	
		n	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**
	600–1190	174	2.42 (0.94)	44	2.55 (0.96)	7	--	490	2.47 (0.92)	86	2.26 (1.24)	288	1.95 (0.95)
	1200–1490	2,189	2.72 (0.83)	762	2.79 (0.95)	270	2.59 (0.99)	5,437	2.96 (0.74)	1,518	2.80 (0.93)	2,629	2.57 (0.84)
	1500–1790	10,034	3.10 (0.72)	4,254	3.17 (0.84)	2,929	2.92 (0.87)	24,341	3.17 (0.68)	10,434	3.12 (0.79)	10,393	2.94 (0.75)
	1800–2090	16,646	3.36 (0.61)	7,566	3.32 (0.74)	8,824	3.16 (0.76)	41,526	3.31 (0.65)	24,911	3.38 (0.69)	15,950	3.25 (0.66)
	2100–2400	5,692	3.51 (0.59)	2,904	3.48 (0.70)	4,161	3.29 (0.71)	13,703	3.45 (0.65)	10,576	3.56 (0.63)	5,367	3.43 (0.65)
Under 50%	Total Group	34,735	3.27 (0.69)	15,530	3.28 (0.79)	16,191	3.14 (0.78)	85,497	3.27 (0.68)	47,525	3.34 (0.73)	34,627	3.12 (0.75)

		Humanities		Mathematics		Natural Sciences		Health Sciences		Social Sciences	
		n	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**	***n***	**M (SD)**
	600–1190	186	2.42 (0.99)	320	1.94 (1.04)	312	1.93 (0.92)	109	2.34 (0.93)	565	2.16 (0.87)
	1200–1490	2,562	2.66 (0.88)	4,810	2.31 (0.99)	4,805	2.29 (0.90)	1,143	2.62 (0.82)	6,674	2.47 (0.80)
	1500–1790	14,679	3.02 (0.75)	24,420	2.65 (0.95)	23,722	2.69 (0.85)	3,889	3.03 (0.78)	29,664	2.84 (0.77)
	1800–2090	28,060	3.34 (0.62)	42,618	2.91 (0.90)	42,768	2.99 (0.79)	4,451	3.40 (0.68)	50,207	3.14 (0.72)
	2100–2400	10,570	3.54 (0.54)	15,113	3.10 (0.89)	15,106	3.17 (0.79)	951	3.61 (0.61)	17,434	3.36 (0.68)
	Total Group	56,057	3.26 (0.70)	87,281	2.83 (0.94)	86,713	2.90 (0.84)	10,543	3.19 (0.79)	104,544	3.05 (0.77)

		Business and Communications		Computer Sciences		Engineering		English		Foreign and Classical Languages		History	
		n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)
50% to 75%	600–1190	5,350	2.50 (1.07)	1,743	2.33 (1.22)	507	2.66 (1.15)	10,234	2.36 (1.07)	1,795	2.46 (1.14)	5,420	1.91 (1.14)
	1200–1490	33,811	2.80 (0.99)	12,115	2.77 (1.15)	5,673	2.82 (1.03)	65,741	2.80 (0.98)	15,973	2.79 (1.00)	36,260	2.36 (1.09)
	1500–1790	62,546	3.13 (0.85)	25,942	3.04 (1.05)	20,819	3.06 (0.92)	124,707	3.19 (0.84)	43,719	3.12 (0.86)	56,739	2.82 (0.97)
	1800–2090	34,446	3.38 (0.72)	20,828	3.15 (0.99)	23,102	3.27 (0.82)	75,324	3.40 (0.73)	39,558	3.36 (0.76)	27,179	3.17 (0.84)
	2100–2400	3,304	3.60 (0.61)	3,982	3.41 (0.86)	4,793	3.50 (0.70)	9,011	3.55 (0.70)	6,020	3.53 (0.76)	2,831	3.45 (0.77)
	Total Group	139,457	3.10 (0.90)	64,610	3.03 (1.07)	54,894	3.16 (0.89)	285,017	3.14 (0.90)	107,065	3.17 (0.88)	128,429	2.74 (1.04)
		Humanities		**Mathematics**		**Natural Sciences**		**Health Sciences**		**Social Sciences**			
		n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)		
	600–1190	3,004	2.18 (1.13)	7,418	1.96 (1.19)	6,835	1.83 (1.11)	2,294	2.49 (1.17)	11,071	2.11 (1.06)		
	1200–1490	22,108	2.62 (1.05)	58,299	2.25 (1.14)	54,066	2.28 (1.05)	15,903	2.94 (1.01)	72,605	2.51 (1.00)		
	1500–1790	53,336	3.00 (0.91)	133,439	2.60 (1.06)	125,160	2.70 (0.95)	27,396	3.22 (0.86)	145,733	2.91 (0.88)		
	1800–2090	36,053	3.28 (0.78)	89,039	2.89 (1.00)	86,862	3.01 (0.87)	14,286	3.41 (0.75)	93,086	3.25 (0.78)		
	2100–2400	4,830	3.54 (0.67)	12,011	3.17 (0.95)	12,289	3.32 (0.80)	2,017	3.60 (0.68)	12,176	3.51 (0.71)		
	Total Group	119,331	3.01 (0.94)	300,206	2.63 (1.09)	285,212	2.72 (0.99)	61,896	3.18 (0.92)	334,671	2.91 (0.94)		

		Business and Communications		Computer Sciences		Engineering		English		Foreign and Classical Languages		History	
		n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)
Over 75%	600–1190	2,783	2.39 (1.09)	781	2.17 (1.32)	73	2.71 (1.18)	4,748	2.43 (1.04)	652	2.30 (1.13)	2,052	1.97 (1.18)
	1200–1490	16,565	2.64 (0.97)	3,195	2.53 (1.25)	933	2.81 (1.08)	25,726	2.81 (0.96)	5,172	2.68 (1.01)	14,073	2.40 (1.08)
	1500–1790	20,820	2.98 (0.89)	3,694	2.85 (1.14)	3,192	3.05 (0.95)	31,112	3.11 (0.89)	9,899	3.05 (0.89)	18,305	2.80 (0.98)
	1800–2090	5,712	3.31 (0.83)	1,519	3.13 (1.05)	2,255	3.26 (0.86)	8,975	3.33 (0.84)	5,141	3.37 (0.80)	5,689	3.13 (0.91)
	2100–2400	287	3.58 (0.71)	137	3.36 (0.98)	299	3.57 (0.69)	592	3.63 (0.66)	532	3.66 (0.68)	393	3.49 (0.73)
	Total Group	46,167	2.87 (0.96)	9,326	2.74 (1.21)	6,752	3.11 (0.95)	71,153	2.99 (0.95)	21,396	3.03 (0.95)	40,512	2.67 (1.06)
		Humanities		Mathematics		Natural Sciences		Health Sciences		Social Sciences			
		n	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)	*n*	M (SD)		
	600–1190	1,339	2.15 (1.12)	1,899	1.81 (1.23)	2,149	1.56 (1.10)	887	2.65 (1.31)	4,644	1.98 (1.08)		
	1200–1490	9,608	2.55 (1.03)	18,343	2.20 (1.20)	17,920	2.12 (1.03)	4,609	2.91 (1.11)	27,228	2.37 (1.00)		
	1500–1790	14,737	2.96 (0.95)	28,759	2.51 (1.17)	28,480	2.56 (0.98)	4,796	3.20 (0.95)	35,946	2.82 (0.93)		
	1800–2090	5,543	3.31 (0.84)	10,664	2.84 (1.09)	10,970	2.98 (0.90)	1,198	3.43 (0.82)	12,161	3.21 (0.84)		
	2100–2400	415	3.59 (0.76)	881	3.26 (0.95)	955	3.39 (0.79)	89	3.76 (0.48)	971	3.59 (0.65)		
	Total Group	31,642	2.87 (1.01)	60,546	2.46 (1.19)	60,474	2.48 (1.05)	11,579	3.07 (1.06)	80,950	2.69 (1.01)		

Note: Means for groups with fewer than 15 students are not shown in the table.

Table 8: Difference Between Average Course Grade and Average Predicted FYGPA

Domain-Specific Area	*n*	Average Course Grade Minus Average Predicted FYGPA
Business and Communications	211,107	0.15
Computer Sciences	85,267	0.04
Engineering	74,395	0.03
English	421,326	0.18
Foreign and Classical Languages	167,185	0.10
History	195,419	-0.10
Humanities	197,087	0.03
Mathematics	428,787	-0.37
Natural Sciences	414,528	-0.29
Health Sciences	80,899	0.23
Social Sciences	497,237	-0.07

Note: A student's high school GPA and scores on each of the three sections of the SAT were used to predict a student's FYGPA within each institution in the sample. Therefore, students without valid HSGPAs are not shown in this table. Positive differences indicate that the domain-specific grade was higher than predicted FYGPA on average. In other words, a student received a higher grade than their predicted FYGPA—the area is graded easier. Negative differences indicate that the domain-specific grade was lower than the predicted FYGPA on average. In other words, a student received a lower grade than their predicted FYGPA—the area is graded harder.

Figure 1: Domain-specific GPAs over time

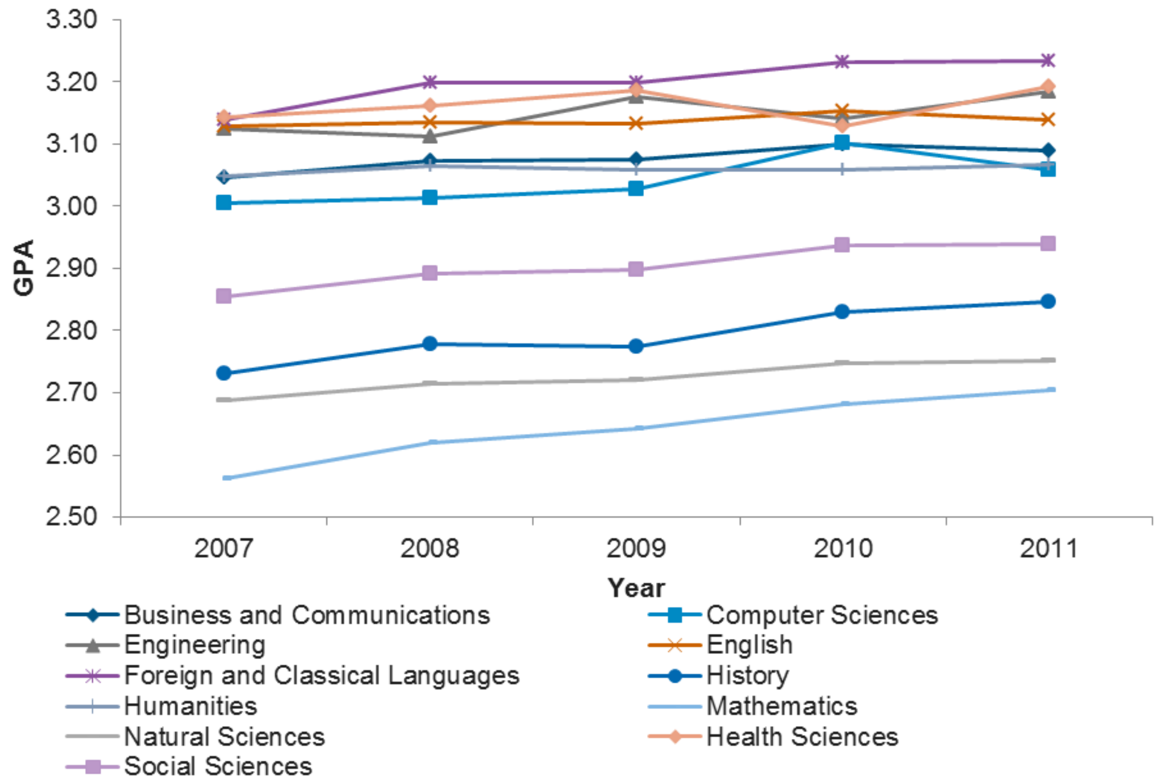


Figure 2: Relationship between average SAT Score and average course grade within a domain for all five cohorts

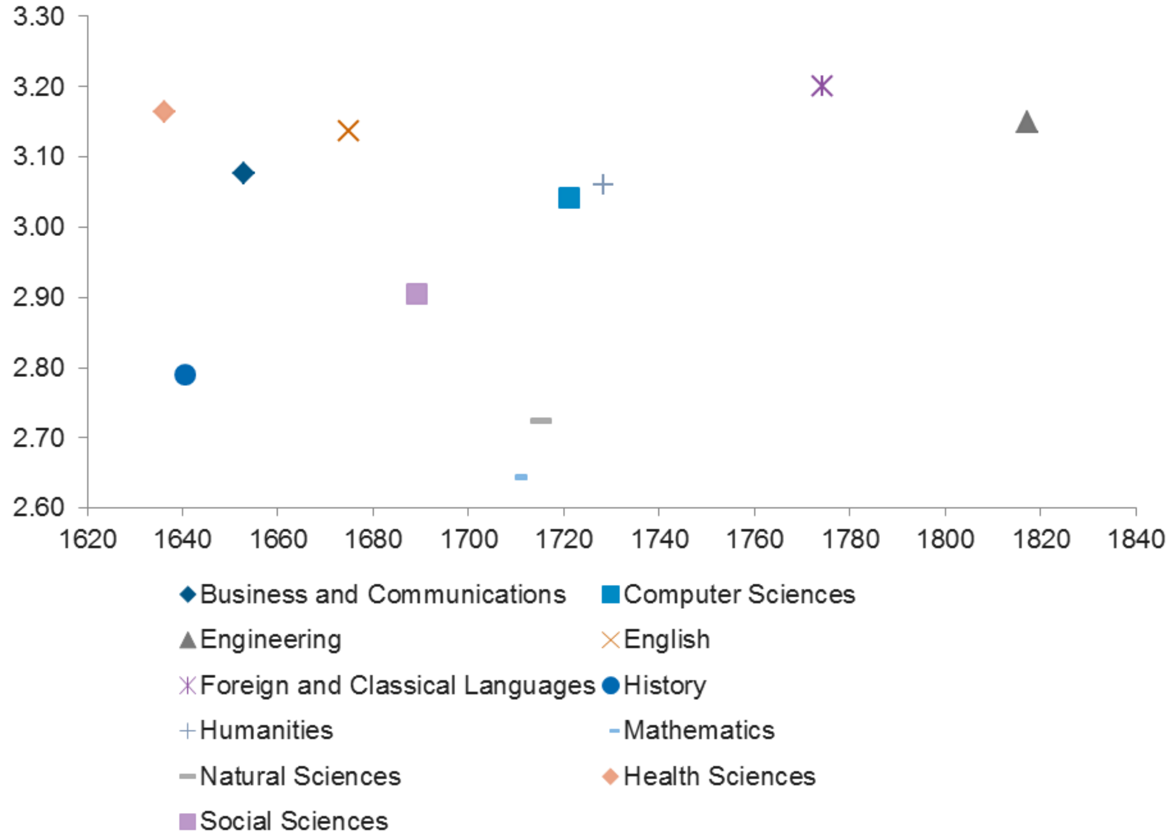
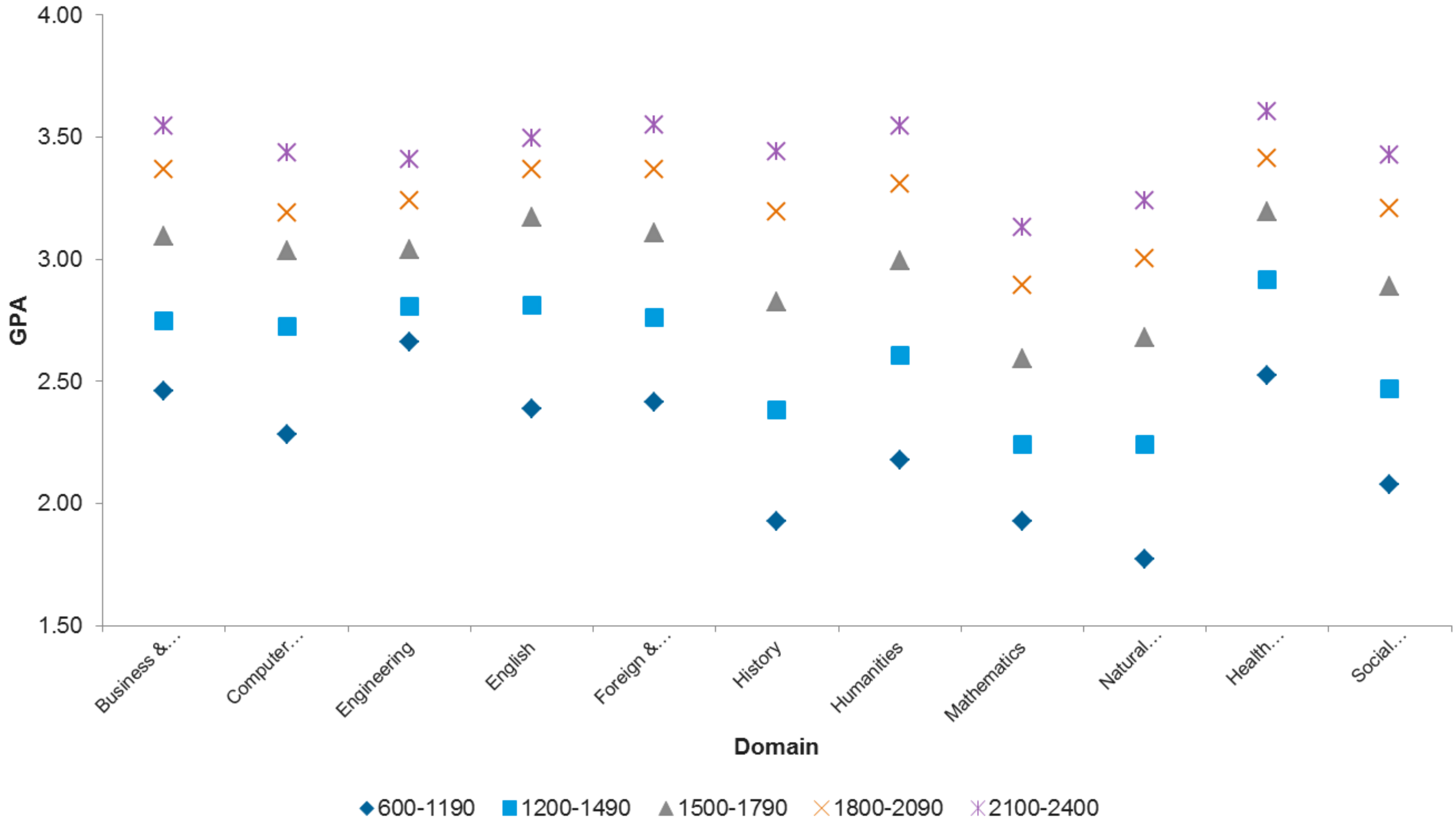


Figure 3: Average domain-specific course grade controlling for SAT score



About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success—including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools. For further information, visit www.collegeboard.org.