ACT Research Report Series 2015 (7)

Differential Grading: Meta-Analyses of STEM and Non-STEM Fields, Gender, and Institutional Admission Selectivity

Paul A. Westrick

Paul Westrick, a research scientist in the Statistical and Applied Research Department at ACT, works on postsecondary outcomes research and validity evidence for the ACT® test and ACT Compass®/ESL.

Acknowledgments

The author thanks Justine Radunzel, Krista Mattern, James Sconing, and Richard Sawyer for their helpful comments and suggestions on earlier drafts of this report.

4685

Contents

Abstract iv
Introduction 1
Differential Grading Research 2
Grading Differences across Institutional Admission Selectivity 2
Grading Differences across Fields of Study 2
Grade Differences between Male and Female Students 3
STEM Fields 4
Ability Levels of STEM Majors 4
STEM Fields: Quantitative and Biological 5
Hypotheses 5
Data 7
Methods 12
Effect Sizes 12
Meta-Analytic Techniques 12
Results 13
Mean ACT Composite Score and HSGPAs Differences between Comparison Groups 13
Mean Undergraduate Semester GPA Differences between SMCs 14
Mean Semester GPA Differences between Male and Female Students 16
Discussion 17
Precollege Academic Achievement 17
Semester GPA, STEM, and Non-STEM Comparisons 18
STEM-Quantitative and STEM-Biological versus STEM 18
Validity Decay and the Dynamic Criterion 19
Gender Comparisons of Semester GPA 19
Hierarchical Moderator Analysis 20
Limitations and Future Research 21
Conclusions 22
References 23
Appendix 27
Tables A-1 to A-11 27

Abstract

This study examined the effects of differential grading in science, technology, engineering, and mathematics (STEM) and non-STEM fields over eight consecutive semesters. Using data from 62,122 students at 26 four-year postsecondary institutions, students were subdivided by institutional admission selectivity levels, gender, and student major categories (SMC): STEM-Quantitative, STEM-Biological, and non-STEM. Estimated mean differences (δ) in semester grade point averages (GPAs) between groups were calculated over eight semesters using meta-analytic techniques.

Within the same gender and institutional admission selectivity level, the average GPA of STEM majors was higher than that of their non-STEM counterparts in the first semester, but average GPA gaps generally decreased over the following semesters. Practically significant differences extended beyond the first year for most of the STEM-Biological versus non-STEM comparisons, however. Institutional admission selectivity and gender moderated the results for the STEM-Quantitative versus non-STEM comparisons and the STEM-Biological versus non-STEM comparisons.

Given the differences in male and female participation rates within the STEM fields, gender comparisons were also conducted. The average GPA of female students was higher than the average GPA of male students, but SMC and admission selectivity moderated the results. The STEM-Biological semester GPA comparisons generally indicated no practically significant differences between male and female students' average GPAs.

Introduction

Grade point average (GPA) is widely used as a measure of academic performance, but questions about the comparability of GPAs have likely been around as long as there have been GPAs. One may ask whether a recent graduate from College A with a GPA of 3.4 is more accomplished than a recent graduate of College B with a GPA of 3.2, or whether a GPA of 3.3 for someone who majored in X is the same as a GPA of 3.3 for someone who majored in Y. In addition to these types of comparisons, differential grading also introduces implications for researchers trying to predict future academic performance. The decline of validity coefficients for precollege academic predictors such as admission test scores and high school grade point averages (HSGPA) over four years of college (Humphreys, 1968; Humphreys & Taber, 1973) may be due to how the criterion—semester GPA—changes with each semester as students specialize within their majors. Differential grading may help explain why validity coefficients decay over time.

Researchers have examined differential grading from multiple perspectives. Some have concentrated their attention on differences between grades earned by students at different types of institutions (Bridgeman, Pollack, & Burton, 2004); by students in different fields of study (e.g., Pennock-Roman, 1994); and by male and female students (e.g., McCormack & McLeod, 1988). Some researchers have examined the differences at a more detailed level—for example, subdividing students by both gender and field of study (Shaw, Kobrin, Patterson, & Mattern, 2012). Collectively, much of this research suggests that grade distributions do differ across fields of study, and that grading standards appear to be more stringent in the science, technology, engineering, and mathematics (STEM) fields. STEM majors tend to enroll at institutions with more-selective admission standards (Chen, 2009), making institutional admission selectivity a factor when examining differential grading. The fact that male and female students are not equally represented in many academic fields, especially the STEM fields (National Science Foundation (NSF), 2013), further complicates comparisons. In light of these issues, this study attempts to provide a better understanding of differential grading by using meta-analysis to estimate mean differences in semester GPA between STEM and non-STEM majors, and between male and female students, with institutional admission selectivity included as a potential moderator.

Differential Grading Research

Grading Differences across Institutional Admission Selectivity

Intuitively, one would expect grading standards to vary across institutions, but the research on differential grading across institutions is sparse, especially in regard to admission selectivity. Thorndike (1963, p. 16) posited that on an absolute scale a "C" at Harvard was higher than an "A" at what he derisively called "Podunk State Teachers College," and Klitgaard (1985) averred that students earning "B" grades at selective schools would most likely earn higher grades if they were enrolled at less-selective schools, but neither provided evidence to support these assertions. More recently, Bridgeman et al. (2004) found that given equal levels of precollege achievement and ability, students attending schools with less-selective admission policies were more likely than students at schools with more-selective admission policies to meet predetermined cumulative GPA benchmarks. Among the students who were in the top level of all three precollege predictor variables (academic intensity, HSGPA, and SAT scores) and enrolled at the most-selective schools, 51% had a freshman GPA of 3.5 or higher. At less-selective schools, 67% of the students in this top-level predictor variable category had a 3.5 GPA or higher, and moving down to the least-selective schools, 72% of the students in this category had a 3.5 GPA or higher.

Grading Differences across Fields of Study

In a follow-up study, Bridgeman, Pollack, and Burton (2008) extended the analyses by looking at the percentage of students earning cumulative GPAs of 3.5 or higher for different types of courses. Students were placed into one of four categories based on their coursework: English, Science/Mathematics/Engineering, Social Sciences, and Education. The results reaffirmed the finding in the previous study that it was more difficult to earn a GPA of 3.5 or higher at the more-selective schools than it was at the less-selective schools. Furthermore, the results indicated that given equal levels of HSGPA, SAT scores, and institutional admission selectivity, it was more difficult to earn a GPA of 3.5 or higher in certain types of courses (e.g., Science, Mathematics/Engineering) than it was in other types of courses (e.g., Education).

Although the Bridgman et al. (2008) study did not provide an in-depth analysis of differential grading practices across fields of study, past research has indicated that grading practices do vary across fields and that grades earned in STEM fields tend to be lower than grades earned in non-STEM fields. In the 1970s, researchers conducted a series of studies at universities in California (Goldman, Schmidt, Hewitt, & Fisher, 1974; Goldman & Hewitt, 1975; Goldman & Widawski, 1976; Hewitt & Jacobs, 1978) that focused on using adaptation-level theory (Helson, 1947, 1948) to understand grading practices. In the Goldman and Widawski (1976) study, the researchers created grading indices for 17 fields of study and found that instructors altered their grading practices according to the ability levels of their students. In general, these studies found that the humanities, social sciences, and other fields had both lower-ability students (based on SAT scores and HSGPA) and lower grading standards than fields in the natural sciences.

During the same decade, other researchers examined the perceived differences in grading practices along with grade inflation. Oh (1976) found that instructors in the natural sciences had the most traditional grading practices, grading students on established standards or on a curve, whereas instructors in the social sciences and humanities were more likely to take student characteristics, such as a disadvantaged background, into account when assigning grades. Furthermore, instructors in the natural sciences were significantly more likely to say that their class activities focused on intellectual or impersonal concepts, whereas instructors in the humanities and social sciences were more likely to report that their class activities were focused on relating knowledge to personal values and attitudes. Prather and Smith (1976, pp. 359–360) found that grading was more stringent in courses "emphasizing quantitative and factual learning" and higher grades were "found in career-oriented courses, such as teacher training." In a subsequent study, Prather, Smith, and Kodras (1979) found that grade inflation was not the result of instructors relaxing their grading standards. Rather, due to proliferation of new, non-traditional majors, students were migrating from the more traditional programs of study "into courses and degree programs which they find have grading standards reflecting their abilities and/or interests" (Prather et al., 1979, p. 23). The general finding from both studies was that there was a strong relationship between fields of study and grading practices, much as Goldman and Widawski (1976) had found.

In response to criticisms of the use of test scores for admission purposes (Crouse, 1985; Owen, 1985), Strenta and Elliott (1987) sought to replicate the Goldman and Widawski (1976) study, creating a grade adjustment index to better compare student performance in different majors at Dartmouth College. In both the Goldman and Widawski (1976) and Strenta and Elliot (1987) studies, students in what are now called STEM fields had mean SAT-Verbal (SAT-V) scores that were roughly equivalent to those of their non-STEM peers, but their mean SAT-Math (SAT-M) scores were higher than those of their non-STEM peers. In a follow-up study, Elliott and Strenta (1988) found that grade adjustments helped reduce validity decay and the under-prediction of female students' academic performance. In both studies, they found their grading indices negatively correlated with SAT scores and high school rank.

Grade Differences between Male and Female Students

Differential prediction for male and female students also served as a catalyst for research on differential grading. Research on the prediction of general academic performance suggested that test scores under-predicted the performance of female students in college (Astin, 1971; Zwick, 2006). That is, when controlling for test scores, female students typically earned higher grades than male students earned. Hewitt and Goldman (1975) concluded that differential course selection almost entirely explained the GPA gap between male and female students at four University of California campuses, and Ramist, Lewis, and McCamley-Jenkins (1994) observed that male students were more likely to take courses with stricter grading standards. McCormack and McLeod (1988) also explored the issue by comparing grades earned by male and female students in 88 individual courses at one university. When looking at cumulative GPA, they found that female students earned higher grades than predicted, but at the individual course level they did not find evidence of gender bias. Pennock-Roman (1994), rather than going down to the individual course level, found the gender gap largely diminished just by controlling for major (discussed later in this report).

One of the more interesting findings by Ramist et al. (1994) was that the gender gaps were much smaller at highly selective schools than those observed at less-selective schools. Zwick (2006) observed that smaller gaps may have resulted from both male and female students at these institutions taking difficult courses. Recent research on differential prediction of second-year cumulative GPA (Shaw et al., 2012) found that even after disaggregating results by academic majors, the use of a common regression line for SAT scores and HSGPA resulted in the underestimation of female students' performance in most of the majors.

STEM Fields

One goal of this study is to determine if grading standards, or, more accurately, grade distributions, in the STEM fields differ from those in the non-STEM fields. In the literature reviewed in the preceding section, the expressions "stricter grading practices" and "stricter grading standards" suggest that students in some majors must have greater levels of achievement to earn the same grade earned by students in other majors. However, it is unclear whether this is always true. It is known that the grade distributions vary across fields of study, but it is not known exactly why they are shifted higher in some fields and shifted lower in others. Furthermore, it is difficult—if not impossible—to argue that achievement on one subject area is the same construct as achievement in another subject area. If the grades earned in subject area X are measures of construct X, and the grades earned in subject area Y are measures of construct Y, grades earned in subject area X cannot mean the same thing as the grades earned in subject area Y.

If achievement in STEM subject areas differs from achievement in non-STEM subject areas, it is necessary to define the STEM fields and demonstrate how they differ from non-STEM fields. Although some consider the social sciences and psychology to be STEM fields (Green, 2007), a more restricted definition (Chen, 2009) limits inclusion to mathematics, natural sciences (including physical sciences and biological/agricultural sciences), engineering/engineering technologies, and computer/information sciences.

These STEM fields have certain characteristics that distinguish them from non-STEM fields. First, STEM fields require students to complete sequential courses in mathematics and the natural sciences (Kokkelenberg & Sinha, 2010; Oh, 1976; Ost, 2010; Prather & Smith, 1976). Elliott and Strenta (1988, p. 334) described mathematics and science curriculums as "hierarchically organized and unforgiving of any lack in basic knowledge or skill." Students cannot opt out of these sequential courses if they want to remain in a STEM field and subsequently earn a STEM degree. Second, past research (discussed above) has suggested that grading standards are more stringent in the STEM fields than the grading standards in other fields.

Ability Levels of STEM Majors

Students seem to be aware that grading standards vary across fields of study (Goldman & Hewitt, 1975; Hewitt & Jacobs, 1978), and they seem to screen themselves into or out of STEM fields on their own. Students enrolling in STEM programs generally have higher levels of precollege academic preparation as measured by HSGPA and admission test scores than do students who enrolled in non-STEM programs (Elliott & Strenta, 1988; Goldman et al., 1974; Goldman & Hewitt, 1975, 1976; Goldman & Widawski, 1976; Green, 1989; Nicholls, Wolfe, Besterfield-Sacre, Shuman, & Larpkiattaworn, 2007; Ost, 2010; Pennock-Roman, 1994; Price, 2010; Strenta & Elliot, 1987; Strenta, Elliot, Adair, Matier, & Scott, 1994; White, 1992), with mathematics scores often identified as the key difference. A National Center for Education Statistics (NCES) study (Chen, 2009) found that although STEM majors made up only 22.8% of the first-year students, they made up 31.1% of the entering students who had earned at least a B average for their HSGPA. Furthermore, they made up 51.1% of the students who had scored in the top quartile on their admission tests. Compared to non-STEM majors, STEM majors were more likely to have enrolled in four-year institutions (51.7% versus 38.1%) and more likely to have enrolled in highly selective institutions (32.6% versus 21.1%).

STEM Fields: Quantitative and Biological

Some researchers have attempted to deal with differential grading practices by developing grade adjustment methods at the individual course level (Berry & Sackett, 2009; Noble & Sawyer, 1987; Stricker, Rock, & Burton, 1993; Young, 1990a, 1990b, 1993), but as Pennock-Roman (1994) observed, such methods were time-consuming. Her alternative was to adjust grades based on majors, creating a scale on which the grades earned by students in given majors were compared to the average grades earned in all courses at a given university. This made it easier to calculate mean differences across majors, but the downside was that the scale differed across institutions. Much like other researchers, Pennock-Roman observed that quantitative skill was the characteristic that separated most majors, something she had found in an earlier study (Pennock-Roman, 1990). However, she also noted that the grading practices in the biological sciences fell somewhere in between those for quantitative and non-quantitative majors. Consequently, she speculated that it might be better to have three groupings for majors: quantitative, non-quantitative, and biological sciences.

The distinction that Pennock-Roman (1994) made between quantitative fields and the biological sciences is important. Grading practices in these two areas may differ, but the distinction is also important because male students make up the majority of students in the quantitative fields while the male-to-female student ratio is more evenly balanced in the biological sciences (NSF, 2013). Female representation in the STEM fields has been an issue of interest for decades (Wai, Cacchio, Putallaz, & Makel, 2010), with researchers (e.g., Ceci & Williams, 2007) providing multiple explanations for why male students are more prevalent than female students are in certain STEM fields, especially the more quantitative fields. Attempting to explain why male and female students enter different STEM fields at different rates is beyond the scope of this paper. However, large differences in male and female participation rates across STEM fields exist. Given these differences in participation rates and that much of the literature on differential grading has emphasized that male-female GPA differences may be due to males and females taking different courses, two STEM categories, STEM-Quantitative and STEM-Biological, will be created in this study.

Hypotheses

Based on the research findings related to STEM majors, differential grading across fields of study, and differences in grades earned by male and female students, a series of hypotheses are proposed. First, as STEM majors enter college with higher mean admission test scores and HSGPAs, STEM majors should be expected to earn higher average grades than those earned by non-STEM majors in the first semester of study, when STEM and non-STEM majors would be expected to take some general education courses together. However, these differences should diminish over time as students take different courses in different majors. This pattern should hold regardless of institutional admission selectivity level and gender.

Hypothesis 1a: Overall, STEM majors will have higher mean semester GPAs than those for non-STEM majors in the first semester of college. In later semesters, the differences between the mean semester GPAs for STEM and non-STEM majors will decrease from those seen in the first semester.

Hypothesis 1b: At both more- and less-selective institutions, STEM majors will have higher mean semester GPAs than those for non-STEM majors in the first semester. In later semesters, the differences between the mean semester GPAs for STEM and non-STEM majors will decrease from those seen in the first semester.

Hypothesis 1c: For both male and female students, STEM majors will have higher mean semester GPAs than those for non-STEM majors in the first semester. In later semesters, the differences between the mean semester GPAs for STEM and non-STEM majors will decrease from those seen in the first semester.

Hypothesis 1d: When disaggregated by institutional admission selectivity levels and gender, STEM majors will have higher mean semester GPAs than those for non-STEM majors in the first semester. In later semesters, the differences between the mean semester GPAs for STEM and non-STEM majors will decrease from those seen in the first semester.

Second, although some researchers have concluded that male GPA and female GPA differences can be explained by differential course-taking (Hewitt & Goldman, 1975; McCormack & McLeod, 1988; Ramist et al., 1994), the most recent—and comprehensive—study (Shaw et al., 2012) suggests that female students should be expected to earn higher grades than male students earn within the same area of study, though these differences may be smaller at more-selective institutions than they are at less-selective institutions (Ramist et al., 1994). Therefore, the following hypotheses are proposed.

Hypothesis 2a: Overall, female students will earn mean semester GPAs that are higher than those for male students in each semester.

Hypothesis 2b: When disaggregated by institutional admission selectivity, female students will earn higher mean semester GPAs than those for male students in each semester, but institutional admission selectivity will moderate the relationship between gender and semester GPA, with the differences being smaller at more-selective institutions than those at less-selective institutions.

Hypothesis 2c: When disaggregated by academic areas of study, female students will earn higher mean semester GPAs than those for male students in each semester.

Hypothesis 2d: When disaggregated by institutional admission selectivity and by academic areas of study, female students will earn higher mean semester GPAs than those for male students in each semester, with the differences being smaller at more-selective institutions than they will be at less-selective institutions.

Data

Starting with an original data set of 209,005 students at 58 four-year institutions that had participated in various ACT research services or partnerships, the final data set consisted of 62,122 students at 26-four-year institutions after screening institutions and students on the following inclusion criteria. A fundamental goal of the study is to compare three student major categories (SMCs)—STEM-Quantitative, STEM-Biological, and non-STEM—within institutions, because institutions that offer all three options give students a choice of entering one of the STEM areas of interest as well as any of the numerous non-STEM fields. Therefore, institutions that did not have at least three observations in each gender by SMC subgroup were excluded. The remaining 26 institutions were located in 13 states, mostly in the Midwest and South of the United States. Twenty-three of the institutions were public and three were private. Admission selectivity was defined in accordance with the ACT Institutional Data Questionnaire, in which institutions self-report their admission selectivity levels. Institutional responses are summarized in Table 1. In this data set, one institution was classified as highly selective; nine were classified as selective; 15 were classified as traditional; and one was classified as open. No institutions in the data set fell into the liberal classification level. As there were not enough institutions at each level, the highly selective and selective institutions were grouped as "more-selective" institutions, and the traditional, liberal, and open admission institutions were grouped as "less-selective" institutions.

Within each institution, student-level inclusion criteria included having valid scores on the ACT® test, HSGPA, semester GPA for each semester, and cumulative GPA for each semester. To ensure that the comparison groups were unchanged across semesters, students had to remain continuously enrolled for eight consecutive semesters in the same four-year institution, from the first semester of the first year through the second semester of the fourth year, or graduate from the initial institution in fewer than eight semesters (n = 3,956).

Table 1. Typical Range of ACT Composite Scores and Class Ranks by Institution Admission Selectivity

Institution Selectivity Level	ACT Composite Scores Middle 50%	Definition
1. Highly Selective	25—30	Majority admitted from top 10% of high school class
2. Selective	21—26	Majority admitted from top 25% of high school class
3. Traditional	18—24	Majority admitted from top 50% of high school class
4. Liberal	17—22	Majority admitted from bottom 50% of high school class
5. Open	16—21	Generally open to all with high school diploma or equivalent

Note: ACT Composite score scale ranges from 1 to 36. Adapted from *National Collegiate Retention and Persistence to Degree Rates* (ACT, 2013).

An additional requirement was that students had classification of instructional program (CIP) codes (NCES, 2002) associated with their records. The CIP code is a six-digit number that identifies a student's declared major. The first two digits are the most general categorization, the first four digits provide an intermediate categorization, and the full six digits provide the most specific categorization. Some majors are more popular than others, and not every major is offered at each institution, so making comparisons at the level of the four-digit or six-digit CIP code was impractical. To maximize sample size, the most general CIP code categorization, two digits, was used. In light of the findings of Pennock-Roman (1994), in this study students with a CIP code of 11 (Computer Sciences), 14 (Engineering), 27 (Mathematics and Statistics), and 40 (Physical Sciences—primarily Physics and Chemistry) were pooled to create the STEM-Quantitative category, and students with a CIP code of 26 (Biological and Biomedical Sciences) were used to create the STEM-Biological category. All other students with a declared major were classified as non-STEM. Students who did not have a valid CIP code for each semester were excluded because they could have been in any of the three categories. Note that the STEM classifications used in this study are a subset of the STEM majors identified in other ACT documents (ACT, 2014).

As grading practices may differ across majors, and these differences affect undergraduate GPA, students who changed from one SMC to another while at the same institution were excluded from the analyses. However, students were allowed to change majors within their SMC. For example, a student who initially majored in mathematics and later changed to engineering would have been included in the study because both majors would fall into the STEM-Quantitative category. If the student had changed from mathematics to communications, the student would have changed categories (STEM-Quantitative to non-STEM) and would have been excluded.

After screening student records on these criteria, a total of 62,122 ACT-tested students who enrolled at one of the 26 institutions as first-time entering college students in fall 2000 through fall 2005 (a total of up to six cohorts per institution) were included in the study. Descriptive statistics for the groups' ACT Composite scores and HSGPAs are presented in Table 2.

Table 2. Means and Standard Deviations for ACT Composite Scores and HSGPA

					ACTC		HSGPA	
Gender	Admission Selectivity	Student Major Category	*k*	*N*	*Mean*	*SD*	*Mean*	*SD*
Male			26	25,055	24.0	4.3	3.53	0.46
Female			26	37,067	23.4	4.0	3.62	0.40
	More		26	44,302	24.4	4.0	3.64	0.39
	Less		26	17,820	21.8	3.9	3.43	0.48
		STEM-Quantitative	26	6,463	26.6	4.1	3.76	0.31
		STEM-Biological	26	3,239	25.6	3.9	3.78	0.30
		Non-STEM	26	52,420	23.1	4.0	3.55	0.44
Male	More		10	18,168	24.8	4.2	3.60	0.42
	Less		16	6,887	21.9	4.0	3.35	0.51
Female	More		10	26,134	24.1	3.9	3.67	0.37
	Less		16	10,933	21.7	3.8	3.49	0.46

Table 2. (continued)

Gender	Admission Selectivity	Student Major Category	*k*	*N*	ACTC *Mean*	ACTC *SD*	HSGPA *Mean*	HSGPA *SD*
Male		STEM-Quantitative	26	4,903	26.5	4.0	3.74	0.33
Male		STEM-Biological	26	1,209	26.0	3.9	3.75	0.32
Male		Non-STEM	26	18,943	23.2	4.1	3.46	0.47
Female		STEM-Quantitative	26	1,560	26.6	4.1	3.83	0.25
Female		STEM-Biological	26	2,030	25.4	3.8	3.79	0.28
Female		Non-STEM	26	33,477	23.1	3.9	3.60	0.41
	More	STEM-Quantitative	10	5,617	26.9	3.9	3.78	0.29
	More	STEM-Biological	10	2,636	26.1	3.6	3.80	0.27
	More	Non-STEM	10	36,049	23.8	3.9	3.61	0.40
	Less	STEM-Quantitative	16	846	24.3	4.0	3.61	0.41
	Less	STEM-Biological	16	603	23.5	4.1	3.67	0.39
	Less	Non-STEM	16	16,371	21.6	3.8	3.42	0.49
Male	More	STEM-Quantitative	10	4,231	26.9	3.9	3.76	0.29
Male	More	STEM-Biological	10	1,015	26.4	3.6	3.77	0.27
Male	More	Non-STEM	10	12,922	23.9	3.9	3.54	0.40
Male	Less	STEM-Quantitative	16	672	24.3	3.8	3.58	0.39
Male	Less	STEM-Biological	16	194	23.7	4.0	3.65	0.40
Male	Less	Non-STEM	16	6,021	21.5	3.7	3.31	0.48
Female	More	STEM-Quantitative	10	1,386	27.0	3.9	3.85	0.22
Female	More	STEM-Biological	10	1,621	25.9	3.5	3.82	0.24
Female	More	Non-STEM	10	23,127	23.8	3.7	3.65	0.34
Female	Less	STEM-Quantitative	16	174	24.2	3.9	3.74	0.33
Female	Less	STEM-Biological	16	409	23.3	3.9	3.68	0.35
Female	Less	Non-STEM	16	10,350	21.6	3.7	3.48	0.44

Note: ACTC = ACT Composite; HSGPA = high school grade point average; *k* = number of institutional studies; STEM = science, technology, engineering, and mathematics.

All institutions reported semester GPA on a four-point scale. For students who graduated prior to their eighth semester, their final cumulative GPAs were used as their semester GPAs for all terms following graduation. Mean semester GPAs over eight semesters are reported for each subgroup in Table A-1 in the appendix. On the right side of each table are two columns, one showing the amount of change in GPA between the first and eighth semesters and the other showing the percent changed. The percent changed is the amount of change before rounding divided by the unrounded GPA for the first semester. For illustrative purposes, Figures 1 through 4 present the mean semester GPAs for the three student major categories (SMCs) disaggregated by gender and institutional admission selectivity levels found in the final twelve rows of Table A-1. While Table A-1 contains the mean semester GPAs for each group across institutions, the differences between groups vary across institutions, hence the need to meta-analyze institutional estimated mean differences and calculate credibility intervals.

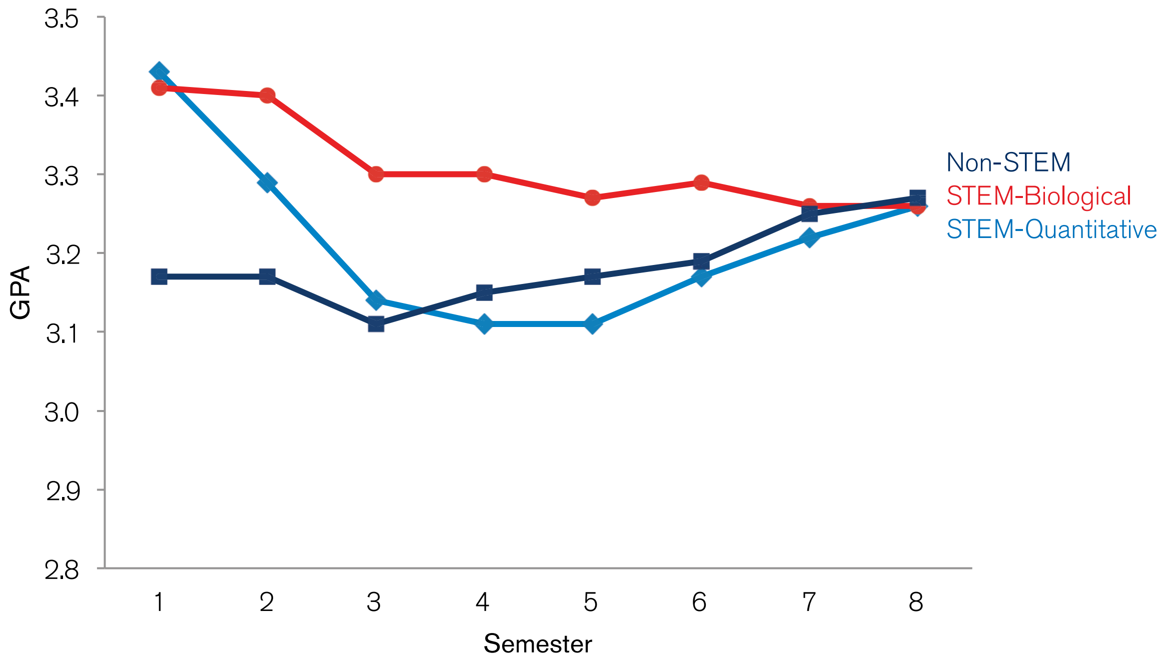

Figure 1. SMC semester GPA trends for female students at more-selective institutions.

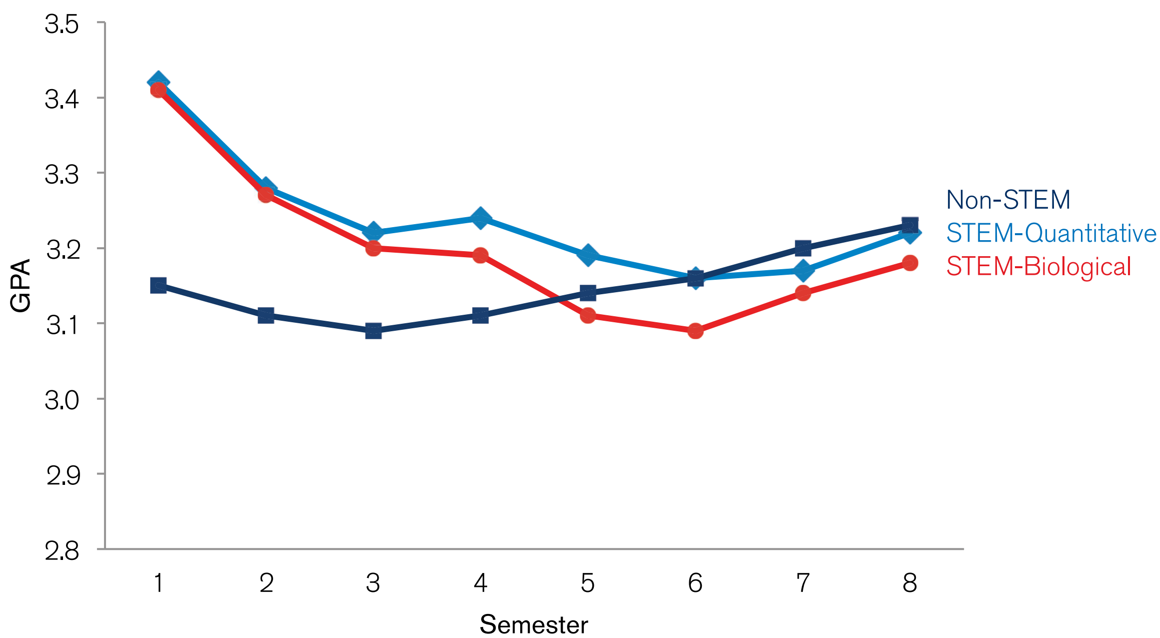

Figure 2. SMC semester GPA trends for female students at less-selective institutions.

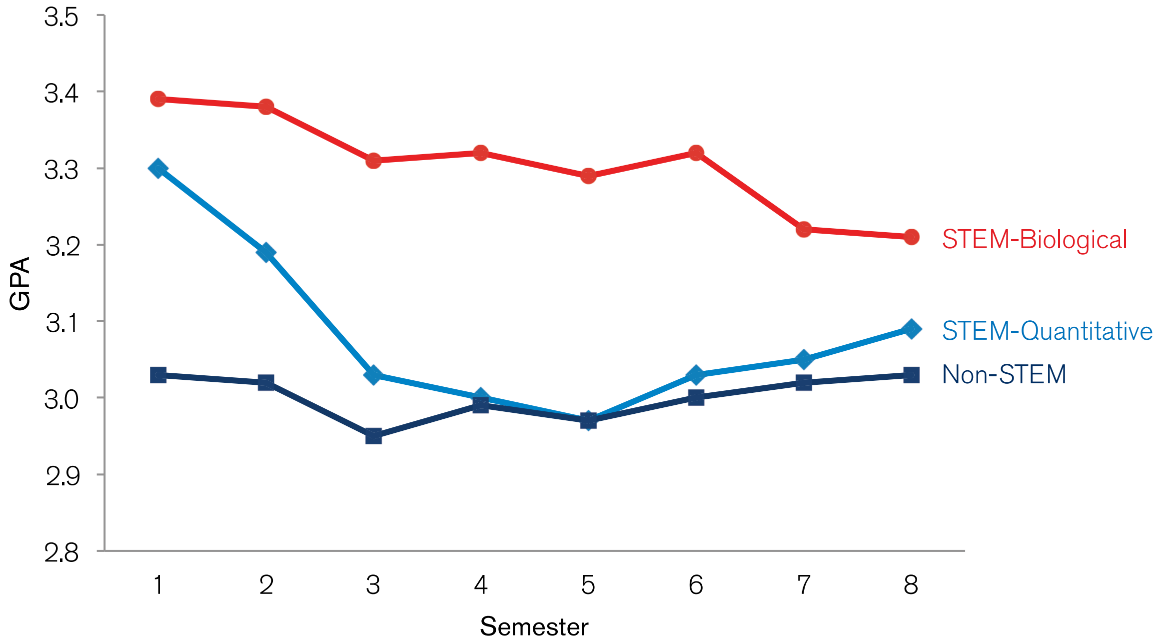

Figure 3. SMC semester GPA trends for male students at more-selective institutions.

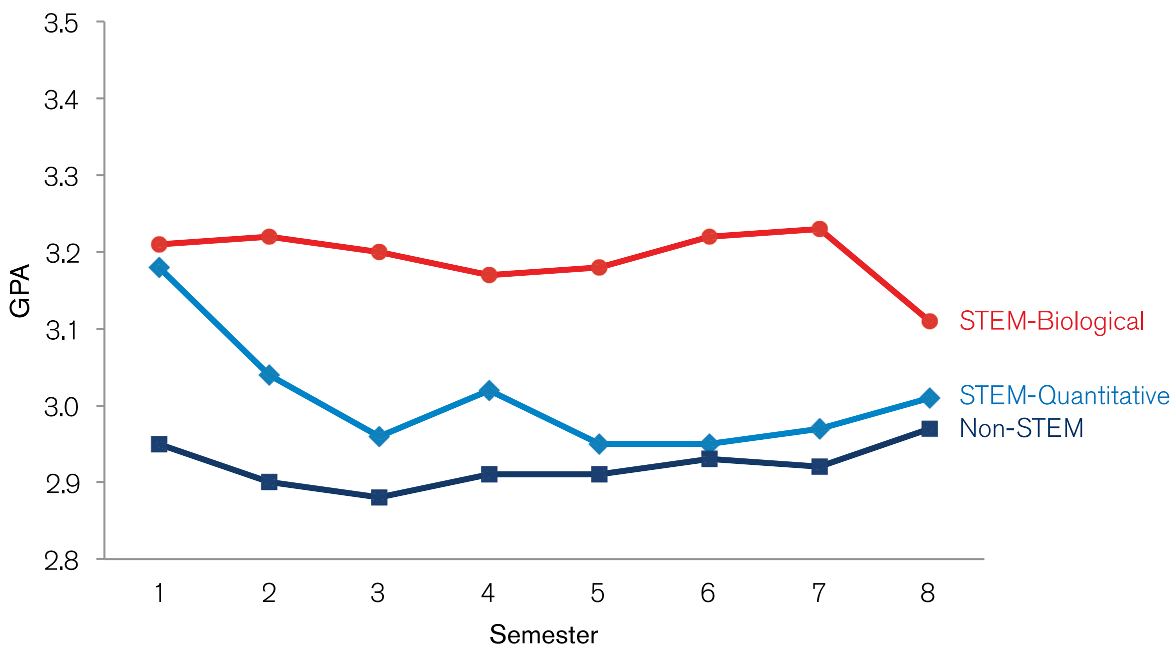

Figure 4. SMC semester GPA trends for male students at less-selective institutions.

Methods

Effect Sizes

The statistic used in this study is Cohen's (1988) *d*, calculated by subtracting the mean of one group from the mean of the comparison group, and then dividing the difference by the pooled standard deviation. According to Cohen's guidelines, effect sizes greater than or equal to |0.20| and less than |0.50| are considered small; *ds* greater than or equal to |0.50| and less than |0.80| are considered medium; and *ds* of |0.80| or greater are considered large.

For the SMC comparisons, at the institution level the mean semester GPAs for the non-STEM majors were subtracted from the mean semester GPAs for the two STEM groups, STEM-Quantitative and STEM-Biological. For the gender comparisons, at the institution level the mean semester GPAs for the females were subtracted from the mean semester GPAs for the males within each SMC. In this study there was no effort to statistically control for differences in precollege academic achievement. Rather, the decision was made to present effect sizes for two precollege measures of academic achievement, ACT Composite scores and HSGPA, and then present effect sizes for semester GPA comparisons. This allowed for comparisons across three types of measures using a common metric.

Meta-Analytic Techniques

Schmidt and Hunter (2014, pp. 279–342) meta-analytic techniques for *d* values were used to analyze the data. For all comparisons, *ds* were calculated at the institution level using the observed ACT Composite scores, HSGPAs, and semester GPAs. However, given that the underlying construct—precollege academic achievement—is measured by both ACT scores and HSGPA, corrections were made for measurement error in each measure. For ACT Composite scores, the reliability estimate was 0.96 (ACT, 2007). For HSGPA, an estimate of 0.79 was used (Schiel & Noble, 1991). The reliability of undergraduate GPA for a full academic year is typically calculated by using the correlation between semester GPAs within an academic year and applying the Spearman-Brown Prophecy Formula. Humphreys (1968) considered the correlations between adjacent semesters to be lower-bound reliability estimates for semester GPA, and as semester GPA is the outcome of interest in this study, the reliability of semester GPA was estimated by using the correlation between adjacent semesters within an academic year.

As gender and institutional admission selectivity are potential moderators for the SMC analyses, the SMCs were further subdivided by these two variables, permitting a hierarchical moderator analysis.[1] Meta-analyses were conducted at the broadest levels (e.g., STEM-Quantitative versus non-STEM), then with one potential moderator (e.g., male, STEM-Quantitative majors versus male, non-STEM majors; and STEM-Quantitative majors at more-selective institutions versus non-STEM majors at more-selective institutions), and then with both potential moderators (e.g., male, STEM-Quantitative majors at more-selective institutions versus male, non-STEM majors at more-selective institutions). For the gender comparisons, the same approach was taken but with SMCs and institutional admission selectivity levels as potential moderators.

[1] In meta-analysis, a hierarchical moderator analysis involves subdividing observations by more than one moderator. As Schmidt and Hunter (2014) describe, conducting moderator analyses one moderator at a time may be misleading. It would be analogous to conducting analysis of variance one variable at a time and not considering interactions. The problem with conducting hierarchical moderator analyses is the need for a sufficient number of studies with observations in each cell for analyses beyond two-way breakouts.

In each set of comparisons, the semester GPA reliability estimates were calculated using only the semester GPAs for the two groups. For example, for the female, STEM-Quantitative versus female, non-STEM comparisons, the correlations between semester GPAs within each academic year at each institution were calculated using only the semester GPAs for the female, STEM-Quantitative and female, non-STEM majors.

In each meta-analysis, a random effects model was used (Schmidt & Hunter, 2014).[2] A random effects model takes into account true differences between institutional studies by allowing for a distribution of effect sizes (in this report, δ) across studies. The random effects model is the recommended approach for meta-analyses (National Research Council, 1992; Schmidt, Oh, & Hayes, 2009). After making corrections for artifacts, the estimated mean effect size (δ) and the standard deviation of the estimated mean effect size ($SD\ \delta$) were calculated. Using the estimated mean effect size and the standard deviation of the estimated mean effect size, 80% credibility intervals were calculated to define the range within which 80% of the true effect sizes would be expected to fall.

The decision to split institutional studies based on institutional admission selectivity levels created the potential for second-order sampling error. Small sample sizes in primary studies create first-order sampling error, and having a small sample of primary studies in a meta-analysis creates second-order sampling error. As with sampling error in primary studies, second-order sampling error affects the standard deviations more than the means (Schmidt & Hunter, 2014). When the sample of studies included in the meta-analysis is small, and many of the studies have effect sizes that are near the estimated mean effect size, the variance left after subtracting the expected variance due to sampling error from the observed variance may be less than zero. While some may find negative variance estimates troubling, Schmidt and Hunter (2014, p. 393) observed that negative variance estimates have also been found in ANOVA and in Generalizability Theory, and that Cronbach, Gleser, Nanda, and Rajaratnam (1972) "recommended substituting 0 for the negative variance, and Brennan (1983) agreed with this recommendation." This recommendation was followed in this study.

Results

Mean ACT Composite Score and HSGPAs Differences between Comparison Groups

Estimated mean differences for ACT Composite scores and HSGPA are presented in Tables A-2 through A-5. Estimated mean effect sizes that exceeded |0.20| and had credibility intervals that did not contain zero are highlighted with bold text. For the eighteen SMC comparisons of ACT Composite scores (Table A-2), overall and within gender and/or admission selectivity levels, all of the effect sizes were positive, indicating that STEM majors had higher mean ACT Composite scores than non-STEM majors. The estimated mean effect sizes were moderate or large, and none of the 80% credibility intervals contained zero. For the twelve gender comparisons of ACT Composite scores (Table A-3), none of the effect sizes for ACT Composite scores exceeded |0.20|, with positive signs indicating that male students had higher scores, and negative signs indicating that female students had higher scores. Most of the credibility intervals contained zero.

[2] As noted in the National Research Council report (1992), the use of the terms "fixed effects" and "random effects" in meta-analysis differs from the use of the terms elsewhere. In meta-analysis, a fixed effects model assumes that population effect sizes in individual studies are homogenous. Given this assumption, there is no variance in population effect sizes across studies, and only a confidence interval is calculated around the population parameter. A random effects model in meta-analysis is somewhat analogous to a random slopes model in hierarchical linear regression.

For the SMC comparisons of average HSGPA (Table A-4), all the effect sizes exceeded 0.20, and none of the credibility intervals contained zero, indicating that the average HSGPAs for the STEM majors were higher than the average HSGPAs for the non-STEM majors. For the STEM-Quantitative and non-STEM comparisons, the estimated mean effect sizes were between 0.47 and 0.48 for the overall and institutional admission selectivity levels, but the effect sizes ranged between 0.58 and 0.64 for all the comparisons that included gender. This was due to male students being more heavily represented in the STEM-Quantitative fields and female students being more heavily represented in the non-STEM fields, as well as female students earning higher HSGPAs in general (see Table 2). For the STEM-Biological comparisons, all the estimated mean effect sizes exceeded 0.20 and were generally larger than those from the STEM-Quantitative versus non-STEM comparisons, except for the three comparisons between female students. None of nine credibility intervals contained zero. The STEM-Biological versus non-STEM results disaggregated by gender indicated that the estimated mean effect sizes were somewhat larger for male students than they were for female students at both more-selective (0.68 vs. 0.54) and less-selective (0.83 vs. 0.55) institutions. Gaps were also seen in the STEM-Biological versus non-STEM ACT Composite score comparisons (Table A-2) at both more- and less-selective institutions, though the 0.28 gap seen in the HSGPA comparisons at less-selective institutions in Table A-4 stood out. As seen in the STEM-Quantitative versus non-STEM comparisons discussed above, comparisons made before disaggregating by gender differ from those made after disaggregating by gender. Whereas the disaggregated results by gender led to larger effect sizes for both male and female analyses for the STEM-Quantitative versus non-STEM comparisons, the STEM-Biological versus non-STEM comparisons made prior to disaggregating by gender masked larger differences seen between the male students and smaller differences seen between the female students.

Gender comparisons of average HSGPA (Table A-5), at the overall and admission selectivity levels suggested that female students have higher HSGPAs than their male counterparts, with all estimated mean effect sizes exceeding -0.20 and none of the credibility intervals containing zero. This pattern held for the STEM-Quantitative and non-STEM majors, but the differences between male and female STEM-Biological majors, though negative, did not exceed -0.20 and the credibility interval contained zero for the more-selective institutions.

In summation, the results in Table A-4 indicate that STEM majors have higher mean HSGPAs than their non-STEM counterparts across institutions. The results in Table A-5 indicate that female students had higher mean HSGPAs than male students among the STEM-Quantitative and non-STEM majors, but the differences were noticeably smaller among the STEM-Biological majors.

Mean Undergraduate Semester GPA Differences between SMCs

Table A-6 contains the effect sizes and credibility intervals for the STEM-Quantitative versus non-STEM comparisons of mean undergraduate semester GPA. Nine sets of comparisons are presented for the overall results and disaggregated by institutional admission selectivity levels and/or gender. The *N* counts are not presented, but they can be found for the corresponding STEM-Quantitative versus non-STEM comparisons made in the upper halves of Tables A-2 (ACT Composite scores) and A-4 (HSGPA). Estimated mean effect sizes that exceeded |0.20| and had credibility intervals that did not contain zero are highlighted with bold text.

Starting with the overall results (upper left corner) in Table A-6, the estimated mean effect size, δ, for the first semester was 0.43, and the 80% credibility interval did not contain zero. This indicated that STEM-Quantitative majors earned higher average semester grades in the first semester than their non-STEM counterparts. In the second semester, the estimated mean effect size exceeded 0.20, but the credibility interval contained zero, indicating that across institutions, STEM-Quantitative majors do not necessarily earn higher average grades in the second semester. From semesters three through eight, all the credibility intervals contain zero, suggesting that there are no differences of practical significance in average grades earned in these semesters by STEM-Quantitative majors and non-STEM majors. The estimated mean effect size declined to 0.03 in the third semester, and the estimated mean effect sizes were negative over the final five semesters, indicating that non-STEM majors earned higher average grades than STEM-Quantitative majors, though none of the effect sizes exceeded -0.20.

The results disaggregated by institutional admission selectivity can be found under the overall results. The general patterns for both more-selective and less-selective institutions were similar to those found in the overall analyses, with the largest estimated mean effect sizes, 0.45 and 0.39, respectively, seen in the first semester, followed by a general decline in estimated mean effect sizes over the following semesters.

The results disaggregated by gender, found to the right of the overall results, also followed the general pattern seen in the overall analyses. In the first semester, the estimated mean effect sizes equaled 0.58 for males and 0.54 for females, and neither credibility interval contained zero. Note that both the estimated mean effect sizes were larger than the 0.43 for the overall results in the first column. The estimated mean effect sizes in the second semester for both male students (0.36) and female students (0.28) exceeded the figure for the overall results (0.21), and, unlike the overall results, neither credibility interval contained zero. None of the estimated mean effect sizes for either group exceeded |0.20| over the following six semesters, and all but one of the credibility intervals contained zero.

For the results disaggregated by institutional admission selectivity and gender, the estimated mean effect sizes were largest in the first semester, ranging between 0.47 and 0.58, and the estimated mean effect sizes diminished over the following semesters. The results more closely paralleled those for the overall gender analyses found in the first row (All Levels: Male, Female) than the results for the overall admission selectivity analyses in the first column (Male and Female, More-Selective, Less-Selective).

Table A-7 contains the estimated mean effect sizes and credibility intervals for the STEM-Biological versus non-STEM comparisons. As in Table A-6, results are presented for the overall analyses and for the analyses disaggregated by institutional admission selectivity and/or gender. The *N* counts can be found for the corresponding STEM-Biological versus non-STEM comparisons made in the lower halves of Tables A-2 and A-4. In the overall analyses, the estimated mean effect size in the first semester was 0.59 and the 80% credibility interval did not contain zero. After the first semester the estimated mean effect sizes declined steadily in each semester, down to 0.06 in the eighth semester. The results disaggregated by admission selectivity and the results disaggregated by gender displayed similar patterns, with first-semester estimated mean effect sizes highest in the first semester and trending downward in later semesters.

For the analyses disaggregated by admission selectivity and gender, the results for the more-selective institutions and females at less-selective institutions displayed a similar pattern, with estimated mean effect sizes ranging between 0.49 and 0.81 in the first semester and then trending downward. The results for the males at less-selective institutions were somewhat inconsistent with the other analyses. Although the estimated mean effect size in the first semester (0.45) exceeded that for the eighth semester (0.23), five of the six estimated mean effect sizes between the first and last semesters exceeded that for the first semester.

In summation, the results for the STEM-Quantitative versus non-STEM comparisons support hypotheses 1a, 1b, 1c, and 1d. However, the results for the STEM-Biological versus non-STEM comparisons supported only hypotheses 1a, 1b, and 1c. Therefore, hypothesis 1d is rejected.

Mean Semester GPA Differences between Male and Female Students

Table A-8 contains the effect sizes and credibility intervals for the gender comparisons made at the overall and admission selectivity levels. All the effect sizes were negative in the overall analyses, meaning that female students earned higher grades than male students, on average, across all eight semesters, with all the estimated mean effect sizes exceeding -0.20 and none of the credibility intervals containing zero. These results support hypothesis 2a. Below the overall analyses are the results disaggregated by institutional admission selectivity. All 16 of the estimated mean effect sizes were negative, and none of the credibility intervals contained zero, providing support for the first part of hypothesis 2b. With the exception of the estimated mean effect size for the first-semester at more-selective institutions (-0.19), all other estimated mean effect sizes exceeded -0.20. The differences also tended to be greater in the later semesters, especially at the more-selective institutions. In each semester, the estimated mean differences were smaller at the more-selective institutions than at the less-selective institutions, fully supporting the second part of hypothesis 2b.

Table A-9 contains the results for the gender comparisons made between male and female students within the same SMC, overall, and disaggregated by institutional admission selectivity levels. Starting with the SMC analyses in the first row, working left to right, among STEM-Quantitative majors, female students earned higher average grades than males in each semester, with seven of the estimated mean effect sizes exceeding -0.20 and none of the eight credibility intervals containing zero. In contrast to the results for the STEM-Quantitative majors, the sign of the estimated mean effect sizes were sometimes positive and sometimes negative for the STEM-Biological majors. None of the estimated mean effect sizes exceeded |0.20|, indicating that there were no differences of practical significance between mean grades earned by male and female STEM-Biological majors. Overall results for the non-STEM majors in the final column of the first row indicated that female students earned higher semester GPAs than male students earned, with estimated mean effect sizes ranging between -0.31 and -0.41, with none of the credibility intervals containing zero. Given the results for the STEM-Biological majors, hypothesis 2c is rejected.

Rows two and three of Table A-9 contain the results for the gender analyses disaggregated by institutional admission selectivity levels and SMCs. The results closely paralleled those found in the SMC analyses in the first row. With the exception of first semester results at more-selective institutions, there were no practical differences between grades earned by male and female

STEM-Biological majors, so the first part of hypothesis 2d was not supported. The second part of hypothesis 2d stated that the effect sizes in each semester at the less-selective institutions would be larger than those in the corresponding semester at the more-selective institutions. This was not true. The differences seen between male and female students were generally, but not always, smaller at the more-selective institutions than they were at the less-selective institutions. As female students did not always have higher mean grades in each semester and the gender differences were not always larger at the less-selective institutions, hypothesis 2d is also rejected.[3]

Discussion

Precollege Academic Achievement

Past research (e.g., Chen, 2009) had found that STEM majors had higher levels of precollege academic achievement, as measured by ACT/SAT scores and HSGPA, than their non-STEM counterparts. Though no hypotheses were made regarding precollege academic achievement, meta-analyses were conducted to examine these differences. The results of these meta-analyses serve as references for the hypotheses for differences between grades earned by STEM and non-STEM majors, and for the differences between grades earned by male and female students. For ACT Composite scores, the estimated mean effect sizes for the STEM versus non-STEM comparisons were moderate to large, and the estimated mean effect sizes for HSGPA were small to large, though moderate to large after disaggregating by gender. The results provide additional evidence that STEM majors enter college with higher levels of precollege academic achievement. While both measures help differentiate between STEM and non-STEM majors, the general trend in the fully disaggregated results was that the effect sizes for ACT Composite scores were larger than the effect sizes for HSGPA, especially in the STEM-Quantitative fields at more-selective institutions.

For the male-female comparisons, there were no differences of practical significance between ACT Composite scores. This was true regardless of institutional admission selectivity level and SMC. The male and female STEM-Biological majors also had no differences of practical significance between their mean HSGPAs, but there were differences of practical significance between male and female HSGPAs among the STEM-Quantitative and non-STEM groups.[4] It's possible that female students in these two SMCs were better prepared for college than their male counterparts, and this advantage in HSGPA carried over into college and was manifested in differences of practical significance in the semester GPA analyses. In contrast, it's also possible that male and female students in the STEM-Biological fields entered college equally prepared, as they had no differences of practical significance regarding their mean ACT Composite scores and HSGPAs, and, given their equal levels of precollege academic performance, this carried over into college where they generally had no differences of practical significance in their mean semester GPAs.

[3] It is worth noting that, on average, the differences were smaller at the more-selective institutions than they were at the less-selective institutions, and this was true for all three of the SMCs. For the STEM-Quantitative majors, the mean δ across eight semesters was -0.22 at the more-selective institutions and -0.33 at the less-selective institutions. For the STEM-Biological majors, the mean δ across eight semesters was -0.04 at the more-selective institutions and -0.07 at the less-selective institutions. For the non-STEM majors, the mean δ across eight semesters was -0.33 at the more-selective institutions and -0.41 at the less-selective institutions.

[4] The results indicated that in their precollege academic achievement levels, differences between STEM and non-STEM majors were greater than those found between males and females. This was true for both ACT Composite scores and HSGPA in the overall analyses and after disaggregating students by institutional admission selectivity.

Semester GPA, STEM, and Non-STEM Comparisons

As discussed above, STEM majors entered college with higher mean ACT scores and HSGPAs than their non-STEM peers. The STEM majors had outperformed the non-STEM majors before entering college, and it was expected that the STEM majors would outperform the non-STEM majors in college, at least in regard to first-semester GPA. After the first semester, the differences in semester GPA were expected to decrease by the eighth semester due to STEM and non-STEM majors taking different courses in different fields of study with different grading practices. Furthermore, this pattern would hold regardless of whether the results were disaggregated by institutional admission selectivity level and/or gender. The results in Tables A-6 and A-7 generally confirmed this expectation, with the only exception being the results for the STEM-Biological versus non-STEM comparisons for males at less-selective institutions. Hypotheses 1a, 1b, and 1c were fully supported, but hypothesis 1d was rejected due to the aforementioned exception.

To highlight the differences across measures, the estimated mean effect sizes for ACT Composite scores, HSGPA, first-semester GPA, and eighth-semester GPA are presented together in Table A-10. The estimated mean effect sizes for first-semester GPA were of practical significance, though generally smaller than those for ACT Composite scores and HSGPA. Note that the estimated mean effect sizes for first-semester GPA most closely paralleled the results for HSGPA. The estimated mean effect sizes for eighth-semester GPA are found in the last column. Whereas all the estimated mean effect sizes in the first semester were positive and of practical significance, nearly half the estimated mean effect sizes were negative in the eighth semester. Of those that exceeded 0.20, all had 80% credibility intervals that contained zero.

The results of these meta-analyses strongly suggest that STEM majors outperformed non-STEM majors before entering college and in the first semester of college when STEM and non-STEM majors likely enroll in general education courses together. The results also suggest that grading distributions, and perhaps grading standards, vary across the three SMCs, as the performance gaps diminish in later semesters when STEM and non-STEM majors are less likely to take courses together.

STEM-Quantitative and STEM-Biological versus STEM

The decision to create STEM-Quantitative and STEM-Biological categories rather than a single STEM group was influenced by Pennock-Roman's (1994) conclusion that the grading practices in the biological sciences did not fit into the grading practices of either the quantitative or the non-quantitative fields. The results of this study provide additional evidence that researchers should separate the two STEM groups. Compared to the STEM-Biological majors, the STEM-Quantitative majors tended to have slightly higher mean ACT Composite scores and HSGPAs. More importantly, there were differences in the estimated mean effect sizes for semester GPA for these two categories when compared to the non-STEM majors. Both STEM groups earned higher average grades than their non-STEM counterparts in the first semester in all the meta-analyses, but the estimated mean effect sizes for the STEM-Quantitative and non-STEM comparisons fell sharply after the first and second semesters. In contrast, the estimated mean effect sizes for the STEM-Biological and non-STEM comparisons declined gradually over time.

For comparative purposes, it is worth noting the semesters when each SMC group achieved its highest and lowest semester GPAs. Referring back to Table A-1, the STEM-Quantitative groups tended to have their highest semester GPAs in the first semester, and their lowest ones between the fourth and sixth semesters, before increasing slightly by the eighth semester. For the STEM-Biological majors, their highest semester GPA was generally in the first semester, and their lowest semester GPA was in the eighth semester. Non-STEM groups had their lowest semester GPA in the third semester, and their highest semester GPA in the eighth semester. At both levels of admission selectivity, the mean GPA for male, non-STEM majors was nearly identical in the first and eighth semesters, but the mean semester GPA for female, non-STEM majors trended upward between the first and eighth semester.

Validity Decay and the Dynamic Criterion

Finally, the results of the SMC comparisons partially explain why validity coefficients reported in a previous study using this data set generally showed less decay when they were calculated for STEM majors only than when they were calculated for STEM and non-STEM majors combined (Westrick, 2012). In the first few semesters, students would be expected to take some general education courses together, and the students with the highest average ACT scores and HSGPAs tended to earn the highest undergraduate grades, resulting in positive correlations between the predictors and the criterion, semester GPA. In later semesters, the validity coefficients generally decreased. It was suggested that in later semesters, students specialized within their areas of study, and if there was little or no overlap of common courses across SMCs, the criterion had changed for each SMC. Hence the use of the term "dynamic criterion" when discussing semester GPA in the later semesters. It was suggested that the STEM majors who had entered college with higher average ACT scores and HSGPAs earned lower grades within their majors. At the same time, the non-STEM majors who had entered college with lower average ACT scores and HSGPAs earned higher grades within their majors. If this were true, it would help explain why the validity coefficients for ACT scores and HSGPA decreased more in the later semesters in the overall analyses than they did in the analyses for STEM majors only.

Evidence of this taking place can be found in Table A-1, which shows that mean semester GPAs for STEM majors tended to decrease over eight semesters, and the mean semester GPAs for non-STEM majors tended to dip slightly over the first two or three years before returning to their first-semester means (analyses for male students) or actually increasing (all other analyses) by the end of the fourth year. Further evidence is the general trend for the standardized differences (δs) between semester GPAs earned by STEM majors and those by non-STEM majors to decrease over time (Tables A-6 and A-7). The results of the current differential grading study helps explain why conducting validity analyses for each of the 12 SMC by gender by institutional admission selectivity groups had reduced the amount of validity decay in the STEM groups, indicating that within each STEM group, students with higher ACT scores and HSGPAs still tended to earn the highest grades through the eighth semester of college.

Gender Comparisons of Semester GPA

The second set of hypotheses stated that female students would earn higher average semester GPAs than male students would earn across eight consecutive semesters, and that this pattern would hold when results were disaggregated by institutional admission selectivity and/or SMC. This study found that female students earned higher undergraduate grades over time than male students in the overall analyses and when disaggregated by institutional admission selectivity level, and the differences were smaller at more-selective institutions, similar to what Ramist et. al. (1994) found.

Following the pattern in Table A-10, estimated mean differences between ACT Composite scores, HSGPA, and semester GPA earned by male and female students are summarized in Table A-11. For semester GPA, the estimated mean effect sizes over eight semesters were averaged. As noted earlier, there were no differences of practical significance for ACT Composite scores, but there were differences of practical significance for HSGPA for all the analyses except for those for the STEM-Biological majors. The results for the semester GPA analyses paralleled those for the HSGPA analyses in that there were differences of practical significance in all the analyses other than those for the STEM-Biological majors. Interestingly, the average estimated mean effect sizes for semester GPA were slightly larger than those for HSGPA in the overall analyses and the analyses disaggregated by institutional admission selectivity. However, after disaggregating the results by the SMCs, the average estimated mean effect sizes for semester GPA were nearly identical to those for HSGPA among the non-STEM majors, and slightly smaller for both STEM groups. Another interesting result is that the average gender differences were generally smaller at the more-selective institutions, though once again the results for the STEM-Biological majors did not follow the trend seen in the other analyses.

Finally, the results of this study may also help explain the differential prediction findings regarding gender that have been reported in another recent ACT study. Radunzel and Noble (2013) found that total-group predictions based on ACT Composite scores generally overestimated long-term college success for male students and underestimated the long-term success of female students. In this study, male students made up the majority of STEM-Quantitative majors and female students made up the majority of non-STEM majors. Although no meta-analyses comparing the mean semester GPAs of male STEM-Quantitative majors and female non-STEM majors were conducted, male STEM-Quantitative majors tended to earn lower average semester GPAs than did female non-STEM majors in the later semesters (Table A-1). Differences in the proportion of male and female students entering the STEM fields may have contributed to the prediction differences found in the study by Radunzel and Noble. However, it is unlikely that differences in the STEM and non-STEM enrollment patterns explain all of the differences in college success rates between male and female students.

Hierarchical Moderator Analysis

Conducting a hierarchical moderator analysis revealed important differences between STEM and non-STEM majors and between males and females that would have been masked if the STEM versus non-STEM analyses had been conducted without considering gender and if the gender analyses had been conducted without considering fields of study. Differences in male and female participation rates in the STEM fields distorted the extent of the differences in semester GPA between STEM and non-STEM majors in the overall and the institutional admission selectivity analyses in Tables A-6 and A-7. Female students in the STEM-Quantitative and non-STEM fields earned higher average semester GPAs than their male counterparts (Table A-9). Consequently, estimated mean effect sizes from the STEM-Quantitative versus non-STEM analyses disaggregated by gender were larger than those in the overall and the institutional admission selectivity analyses (Table A-6). In contrast, the differences between the average semester GPAs earned by male and female students in the STEM-Biological fields were minimal (Table A-9). Given the semester GPA gaps between male and female students in the non-STEM fields, the estimated mean effect sizes from the STEM-Biological versus non-STEM analyses for the male students were larger—and those for the female students smaller—than those from the overall and the institutional admission selectivity analyses (Table A-7).

The hierarchical moderator analyses also provided insights on differences between semester GPAs earned by male and female students (Tables A-8 and A-9). The overall and institutional admission selectivity analyses would suggest that female students earn higher semester GPAs than their male counterparts, and that the differences are even greater than those seen between their mean HSGPAs (Table A-5). However, differences in male and female participation rates in the STEM fields distort these results. After disaggregating the data by the three SMCs, the semester GPA differences for the STEM majors were actually smaller than those in the overall and institutional admission selectivity analyses and smaller than the corresponding differences in HSGPA for each STEM comparison group.

Limitations and Future Research

The results of this study provide valuable insights on grading differences between STEM and non-STEM fields and on gender differences in grades earned within the same general areas of study. This study, however, is not without limitations. Foremost among these was the study's descriptive nature. Estimated mean differences were calculated between comparison groups within each institution, and the results were aggregated across institutions in the meta-analyses. The benefit of this approach was the simplicity of estimating the effect sizes and the comparisons made between the effect sizes for different analyses. However, the results do not explain the extent to which differences in mean semester GPAs can be explained by differences in precollege academic achievement. Statistically controlling for ACT scores and HSGPA in hierarchical regression analyses would better answer questions of that nature.

Another limitation was that students had to be continuously enrolled over eight consecutive semesters. This made the results easier to interpret as the same students were included in the analyses over time. However, conducting the analyses to include all students enrolled in each semester probably would have produced different results in semesters one through seven. Students who drop out of college generally earn lower grades than students who persist in college, and including the students who would later drop out of college would have lowered the mean semester GPAs for all groups.

Another methodological decision that influenced the results was the pooling of academic majors within the SMCs. This pooling may have hidden important distinctions between majors within each SMC. A finer analysis of the data at the two-digit CIP code in the STEM-Quantitative and non-STEM SMCs may provide important insights, such as smaller gender differences in semester GPA. Unfortunately, small sample sizes at many of the institutions made this impractical for the STEM-Quantitative majors and many of the smaller non-STEM majors. Given large enough sample sizes, more two-digit CIP families could be studied and analyses could even be conducted at the four- or six-digit CIP code level for some majors.

Past research on grades earned by male and female students had suggested that female students earned higher grades because male and female students often took different courses in different fields of study, with male students more likely to take courses with rigorous grading standards (e.g., Hewitt & Goldman, 1975; Ramist et al., 1994), such as those found in the quantitative fields. While the results of the this study appear to suggest otherwise, it is important to acknowledge that the SMCs contained multiple fields of academic study, and male and female students may have been enrolled in different programs and different courses within each SMC. This study did not look at grades earned in individual courses, the approach taken by other researchers (e.g., McCormack &

McLeod, 1988). It may have been that within each SMC, male and female students did take different courses. Future research should evaluate course-level data from multiple institutions to further investigate this question.

Another limitation of this study was the absence of any psychosocial measures. Cognitive assessments such as the ACT are highly correlated with HSGPA and undergraduate GPA (ACT, 2007). Psychosocial measures are also positively correlated with high school grades (Allen & Robbins, 2010), but they are not highly correlated with cognitive measures. Including psychosocial measures in future studies may help explain the gender differences in HSGPA and undergraduate GPA.

Finally, second-order sampling error appeared to be an issue, especially for the gender comparisons for STEM-Quantitative and STEM-Biological majors. Admission selectivity had been included as a potential moderator because past research had found gender differences were smaller at more-selective institutions than they were at less-selective institutions (Ramist et al., 1994). Study results confirmed previous research on this topic. The meta-analyses conducted without admission selectivity as a moderator helped to reduce the amount of second-order sampling error in many, but not all, cases. The effect sizes for the overall analyses, however, closely paralleled the results for the meta-analyses conducted at the more-selective institutions because the more-selective institutions had larger sample sizes, masking the differences in effect sizes that emerged between the more- and less-selective institutions. Therefore, admission selectivity should be included as a moderator in future studies, but it would be desirable to have more institutional studies at each admission selectivity level to avoid issues with second-order sampling error.

Conclusions

This meta-analysis has made important contributions to the literature on differential grading. Using data from 62,122 students at 26 four-year institutions, the results strongly suggest that STEM majors earn higher grades in the first semester of college, when STEM and non-STEM majors alike are typically together in general education courses. As students specialize within their majors, the semester GPAs for the STEM majors tend to decline over the following seven semesters, but the semester GPAs for the non-STEM majors tend to remain level (for male students) or increase (for female students) over the same time period. By the eighth semester, the mean differences in semester GPA largely disappear. It would be misleading, however, to interpret the eighth semester GPAs of STEM and non-STEM majors as measures of the same criterion. The results of this study do not prove that different fields have different grading standards, but the results do indicate that semester GPA trends tend to vary across SMCs. When conducting research on semester GPA (and cumulative GPA) trends, ignoring fields of study may lead to erroneous conclusions.

This study has also made contributions to the research on gender differences in grades earned. Gender comparisons of semester GPA that ignore precollege academic achievement levels and undergraduate fields of study may be misleading. Overall and within institutional admission selectivity grouping, the results indicate that among the STEM-Quantitative and non-STEM groupings, female students earn higher average grades than their male counterparts. This may be a carryover from practical differences in their mean HSGPAs. In contrast, male and female STEM-Biological majors have no differences of practical significance in their mean ACT Composite scores and HSGPAs, and with only one exception, they have no differences of practical significance in their undergraduate semester GPAs over four years. ■

References

ACT. (2007). *The ACT technical manual*. Iowa City, IA: ACT.

ACT. (2013). *National collegiate retention and persistence to degree rates*. Iowa City, IA: ACT.

ACT. (2014). *The Condition of STEM 2014*. Iowa City, IA: ACT.

Allen, J. & Robbins, S. (2010). Effects of interest-major congruence, motivation, and academic performance on timely degree attainment. *Journal of Counseling Psychology, 57*(1), 23–35.

Astin, A. W. (1971). *Predicting academic performance in college*. New York, NY: The Free Press.

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

Brennan, R. L. (1983). Elements of generalizability theory. Iowa City, IA: ACT.

Bridgeman, B., Pollack, J., & Burton, N. (2004). *Understanding what SAT Reasoning test scores add to high school grades: A straightforward approach* (College Board Research Report No. 2004-04). New York, NY: The College Board.

Bridgeman, B., Pollack, J., & Burton, N. (2008). *Predicting grades in different types of college courses* (College Board Research Report 2008-1, ETS RR-08-06). New York, NY: The College Board.

Ceci, S. J. (Ed), & Williams, W. M. (Ed). (2007). *Why aren't more women in science: Top researchers debate the evidence*. Washington, DC: American Psychological Association.

Chen, X. (2009). *Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education* (NCES 2009-161). National Center for Education Statistics. Washington, DC: US Department of Education.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Crouse, J. (1985). Does the SAT help colleges make better selection decisions? *Harvard Educational Review, 55*, 195–219.

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

Goldman, R. D., & Hewitt, B. N. (1975). Adaption-level as an explanation for differential standards in college grading. *Journal of Educational Measurement, 12,* 149–161.

Goldman, R. D., & Hewitt, B. N. (1976). The Scholastic Aptitude Test "explains" why college men major in Science more often than college women. *Journal of Counseling Psychology, 23,* 50–54.

Goldman, R. D., Schmidt, D. E., Hewitt, B. N., & Fisher, R. (1974). Grading practices in different major fields. *American Education Research Journal, 11*(4), 343–357.

Goldman, R. D., & Widawski, M. H. (1976). A within-subjects technique for comparing college grading standards: Implications in the validity of the evaluation of college achievement. *Educational and Psychological Measurement, 36*, 381–390.

Green, K. C. (1989). A profile of undergraduates in the sciences. *American Scientist, 77*, 475–480.

Green, M. (2007). *Science and Engineering Degrees: 1966–2004 (NSF 07-307).* Arlington, VA: National Science Foundation.

Helson, H. (1947). Adaption-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology, 60,* 1–29.

Helson, H. (1948). Adaption-level as a basis for a quantitative theory of frame of reference. *Psychological Review, 55,* 297–313.

Hewitt, B. N., & Goldman, R. D. (1975). Occam's razor slices through the myth that college women overachieve. *Journal of Educational Psychology, 67*(2), 325–330.

Hewitt, B. N., & Jacobs, R. (1978). Student perceptions of grading practices in different major fields. *Journal of Educational Measurement, 15,* 213–218.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology, 59*, 375–380.

Humphreys, L. G., & Taber, T. (1973). Postdiction study of the GRE and eight semesters of college grades. *Journal of Educational Measurement, 10*, 179–184.

Klitgaard, R. (1985). *Choosing elites*. New York, NY: Basic Books.

Kokkelenberg, E. C., & Sinha, E. (2010). Who succeeds in STEM studies? An analysis of Binghamton University undergraduate students. *Economics of Education Review, 29,* 935–946.

McCormack, R. L., & McLeod, M. M. (1988). Gender bias in the prediction of college course performance. *Journal of Educational Measurement, 25,* 321–331.

National Center for Education Statistics. (2002). Classification of Instructional Programs—2000. (2002). *US Department of Education National Center for Education Statistics* (NCES 2002-165). Washington, DC: US Government Printing Office.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy of Science Press.

National Science Foundation, National Center for Science and Engineering Statistics. (2013). *Science and Engineering Degrees: 1966–2010*. Detailed Statistical Tables NSF 13-327. Arlington, VA.

Nicholls, G. M., Wolfe, H., Besterfield-Sacre, M., Shuman, L. J., & Larpkiattaworn, S. (2007). A method for identifying variables for predicting STEM enrollment. *Journal of Engineering Education, 96*, 33–43.

Noble, J., & Sawyer, R. (1987). *Predicting grades in specific college freshman courses from ACT test scores and self-reported high school grades* (ACT Research Report No. 87-20). Iowa City, IA: ACT.

Oh, E. (1976). *A study of instructor grading philosophies and practices in undergraduate courses at Western Michigan University*. (Doctoral dissertation). Retrieved from Dissertation and Thesis database. (UMI No. 302847348)

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review, 29*, 923–934.

Owen, D. (1985). *None of the above: Behind the myth of scholastic aptitude*. Boston: Houghton Mifflin.

Pennock-Roman, M. (1990). *Test validity and language background: A study of Hispanic American students at six universities*. New York: The College Board.

Pennock-Roman, M. (1994). *College major and gender differences in the prediction of college grades* (College Board Report No. 94-2). New York, NY: The College Board.

Prather, J. E., & Smith, G. (1976). A study of the relationships between faculty characteristics, subject field, and course grading patterns. *Research in Higher Education, 5,* 351–363.

Prather, J. E., Smith, G., & Kodras, J. E. (1979). A longitudinal study of grades in 144 undergraduate courses. *Research in Higher Education, 10,* 11–24.

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review, 29*, 901–910.

Radunzel, J., & Noble, J. (2013). *Differential effects on student demographic groups of using ACT College Readiness Assessment Composite score, ACT benchmarks, and high school grade point average for predicting long-term college success through degree completion* (ACT Research Report 2013-5). Iowa City, IA: ACT.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting grades: Sex, language, and ethnic groups.* (College Board Research Report 93-1). New York, NY: The College Board.

Schiel, J., & Noble, J. (1991). *Accuracy of self-reported course work and grade information of high school sophomores* (ACT Research Report 91-6). Iowa City, IA: ACT.

Schmidt, F. L. & Hunter, J. E. (2014). *Methods of meta-analysis* (3rd Ed.). Thousand Oaks, CA: Sage.

Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 97–128.

Shaw, E. J., Kobrin, J. L., Patterson, B. F., & Mattern, K. D. (2012). *The validity of the SAT for predicting cumulative grade point average by college major*. (College Board Research Report No. 2012-6). New York, NY: The College Board.

Strenta, A. C., & Elliot, R. (1987). Differential grading revisited. *Journal of Educational Measurement, 24*(4), 281–291.

Strenta, A. C., Elliot, R., Adair, R., Matier, M., & Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Research in Higher Education, 35*(5), 513–547.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85*(4), 710–718.

Thorndike, R. L. (1963). *The concepts of over- and underachievement*. New York, NY: Bureau of Publication, Teachers College, Columbia University.

Wai, J., Cacchio, M., Putallaz, M., & Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30-year examination. *Intelligence 38*, 412–423.

Westrick, P. A. (2012). *Validity decay versus validity stability in STEM and non-STEM fields*. (Doctoral dissertation, University of Iowa). Retrieved from ir.uiowa.edu/etd/3402/

White, P. E. (1992). *Women and minorities in science and engineering: An update* (NSF-92-303). Washington, DC: National Science Foundation.

Young, J. W. (1990a). Adjusting the cumulative GPA using Item Response Theory. *Journal of Educational Measurement, 27,* 175–186.

Young, J. W. (1990b). Are validity coefficients understated due to correctable defects in the GPA? *Research in Higher Education, 31,* 319–325.

Young, J. W. (1993). Grade adjustment methods. *Review of Educational Research, 63*, 151–165.

Zwick, R. (2006). Higher education admission testing. In R. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 647–679). Westport, CT: American Council on Education, Praeger.

Appendix

Tables A-1 to A-11

Table A-1. Mean Semester Grade Point Averages, by Gender, Admission Selectivity, and Student Major Category

Gender	Admission Selectivity	Student Major Category	*k*	*N*	Semester 1	2	3	4	5	6	7	8	Change 1–8	Percent Change
Male			26	25,055	**3.08**	3.04	***2.96***	2.99	2.97	3.00	3.01	3.03	-0.05	-1.6%
Female			26	37,067	3.19	3.17	***3.11***	3.15	3.16	3.18	3.24	**3.26**	+0.07	+2.2%
	More		10	44,302	3.16	3.14	***3.07***	3.10	3.10	3.13	3.16	**3.18**	+0.02	+0.6%
	Less		16	17,820	3.10	3.05	***3.02***	3.04	3.05	3.07	3.10	**3.13**	+0.03	+1.0%
		STEM-Quantitative	26	6,463	**3.32**	3.20	3.05	3.03	***3.00***	3.05	3.08	3.12	-0.20	-6.0%
		STEM-Biological	26	3,239	**3.39**	3.37	3.28	3.29	3.25	3.27	3.23	***3.22***	-0.17	-5.0%
		Non-STEM	26	52,420	3.11	3.09	***3.04***	3.08	3.08	3.11	3.15	**3.17**	+0.06	+1.9%
Male	More		10	18,168	**3.11**	3.08	2.99	3.01	***2.99***	3.03	3.04	3.05	-0.06	-1.9%
Male	Less		16	6,887	**2.98**	2.93	***2.90***	2.93	2.92	2.94	2.94	2.97	-0.01	-0.3%
Female	More		10	26,134	3.20	3.19	***3.12***	3.16	3.17	3.20	3.25	**3.27**	+0.07	+2.2%
Female	Less		16	10,933	3.17	3.12	***3.10***	3.11	3.14	3.15	3.20	**3.23**	+0.06	+1.9%
Male		STEM-Quantitative	26	4,903	**3.28**	3.17	3.02	3.00	***2.96***	3.02	3.04	3.08	-0.20	-6.1%
Male		STEM-Biological	26	1,209	**3.36**	3.35	3.29	3.30	3.27	3.30	3.22	***3.19***	-0.17	-5.1%
Male		Non-STEM	26	18,943	3.00	2.98	***2.93***	2.97	2.95	2.98	2.99	**3.01**	+0.01	+0.3%
Female		STEM-Quantitative	26	1,506	**3.42**	3.28	3.15	***3.12***	***3.12***	3.17	3.22	3.25	-0.17	-5.0%
Female		STEM-Biological	26	2,030	**3.41**	3.37	3.28	3.28	3.24	3.25	***3.23***	3.24	-0.17	-5.0%
Female		Non-STEM	26	33,477	3.17	3.15	***3.10***	3.14	3.16	3.18	3.24	**3.26**	+0.09	+2.8%
	More	STEM-Quantitative	10	5,617	**3.33**	3.21	3.06	3.03	***3.00***	3.06	3.09	3.13	-0.20	-6.0%
	More	STEM-Biological	10	2,636	**3.40**	3.39	3.30	3.31	3.28	3.30	3.24	***3.24***	-0.16	-4.7%
	More	Non-STEM	10	36,049	3.12	3.11	***3.05***	3.09	3.10	3.12	3.17	**3.19**	+0.07	+2.2%
	Less	STEM-Quantitative	16	846	**3.23**	3.09	3.01	3.06	3.00	***3.00***	3.01	3.05	-0.18	-5.6%
	Less	STEM-Biological	16	603	**3.34**	3.26	3.20	3.19	3.13	***3.13***	3.17	3.15	-0.19	-5.7%
	Less	Non-STEM	16	16,371	3.08	3.04	***3.01***	3.04	3.05	3.07	3.10	**3.13**	+0.05	+1.6%
Male	More	STEM-Quantitative	10	4,231	**3.30**	3.19	3.03	3.00	***2.97***	3.03	3.05	3.09	-0.21	-6.2%
Male	More	STEM-Biological	10	1,015	**3.39**	3.38	3.31	3.32	3.29	3.32	3.22	***3.21***	-0.18	-5.3%
Male	More	Non-STEM	10	12,922	3.03	3.02	***2.95***	2.99	2.97	3.00	3.02	**3.03**	+0.00	+0.0%

Note: k = number of institutional studies; **bold** = highest semester GPA for group before rounding; ***bold italics*** = lowest semester GPA for group before rounding.

Table A-1. (continued)

Gender	Admission Selectivity	Student Major Category	*k*	*N*	Semester 1	2	3	4	5	6	7	8	Change 1–8	Percent Change
Male	Less	STEM-Quantitative	16	672	**3.18**	3.04	2.96	3.02	***2.95***	2.95	2.97	3.01	-0.17	-5.4%
Male	Less	STEM-Biological	16	194	3.21	3.22	3.20	3.17	3.18	3.22	**3.23**	***3.11***	-0.10	-3.2%
Male	Less	Non-STEM	16	6,021	2.95	2.90	***2.88***	2.91	2.91	2.93	2.92	**2.97**	+0.01	+0.4%
Female	More	STEM-Quantitative	10	1,386	**3.43**	3.29	3.14	***3.11***	3.11	3.17	3.22	3.26	-0.17	-4.9%
Female	More	STEM-Biological	10	1,621	**3.41**	3.40	3.30	3.30	3.27	3.29	3.26	***3.26***	-0.14	-4.2%
Female	More	Non-STEM	10	23,127	3.17	3.17	***3.11***	3.15	3.17	3.19	3.25	**3.27**	+0.10	+3.1%
Female	Less	STEM-Quantitative	16	174	**3.42**	3.28	3.22	3.24	3.19	***3.16***	3.17	3.22	-0.20	-5.9%
Female	Less	STEM-Biological	16	409	**3.41**	3.27	3.20	3.19	3.11	***3.09***	3.14	3.18	-0.23	-6.8%
Female	Less	Non-STEM	16	10,350	3.15	3.11	***3.09***	3.11	3.14	3.16	3.20	**3.23**	+0.08	+2.5%

Note: k = number of institutional studies; **bold** = highest semester GPA for group before rounding; ***bold italics*** = lowest semester GPA for group before rounding.

Table A-2. Estimated Mean ACT Composite Score Differences SMC Comparisons

Group 1			Group 2							
Gender	Admission Selectivity	SMC	Gender	Admission Selectivity	SMC	*k*	*N*	*d*	*δ*	80% CrI
		Quantitative			Non-STEM	26	58,883	0.74	**0.77**	[0.47, 1.07]
Male		Quantitative	Male		Non-STEM	26	23,846	0.72	**0.74**	[0.41, 1.08]
Female		Quantitative	Female		Non-STEM	26	35,037	0.77	**0.79**	[0.48, 1.11]
	More	Quantitative		More	Non-STEM	10	41,666	0.78	**0.81**	[0.50, 1.12]
	Less	Quantitative		Less	Non-STEM	16	17,217	0.66	**0.68**	[0.45, 0.91]
Male	More	Quantitative	Male	More	Non-STEM	10	17,153	0.73	**0.75**	[0.41, 1.10]
Female	More	Quantitative	Female	More	Non-STEM	10	24,513	0.82	**0.85**	[0.52, 1.18]
Male	Less	Quantitative	Male	Less	Non-STEM	16	6,693	0.69	**0.72**	[0.42, 1.01]
Female	Less	Quantitative	Female	Less	Non-STEM	16	10,524	0.64	**0.66**	[0.47, 0.85]
		Biological			Non-STEM	26	55,659	0.63	**0.66**	[0.36, 0.95]
Male		Biological	Male		Non-STEM	26	20,152	0.71	**0.73**	[0.42, 1.04]
Female		Biological	Female		Non-STEM	26	35,507	0.59	**0.61**	[0.35, 0.88]
	More	Biological		More	Non-STEM	10	36,685	0.66	**0.68**	[0.36, 1.00]
	Less	Biological		Less	Non-STEM	16	16,974	0.58	**0.60**	[0.37, 0.82]
Male	More	Biological	Male	More	Non-STEM	10	13,937	0.73	**0.76**	[0.38, 1.13]
Female	More	Biological	Female	More	Non-STEM	10	24,748	0.61	**0.64**	[0.34, 0.93]
Male	Less	Biological	Male	Less	Non-STEM	16	6,215	0.65	**0.67**	[0.67, 0.67]
Female	Less	Biological	Female	Less	Non-STEM	16	10,759	0.54	**0.56**	[0.37, 0.75]

Note: SMC = student major category; *k* = number of institutional studies; *d* = mean observed effect size; *δ* = estimated mean effect size after corrections for measurement error; CrI = credibility interval; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero.

Table A-3. Estimated Mean ACT Composite Score Differences for Gender Comparisons

Group 1			Group 2							
Gender	Admission Selectivity	SMC	Gender	Admission Selectivity	SMC	*k*	*N*	*d*	δ	80% CrI
Male			Female			26	62,122	0.13	0.14	[-0.02, 0.29]
Male	More		Female	More		10	44,302	0.17	0.18	[0.07, 0.29]
Male	Less		Female	Less		16	17,820	0.03	0.03	[-0.13, 0.20]
Male		Quantitative	Female		Quantitative	26	6,463	-0.02	-0.02	[-0.04, 0.00]
Male	More	Quantitative	Female	More	Quantitative	10	5,716	-0.01	-0.01	[-0.06, 0.04]
Male	Less	Quantitative	Female	Less	Quantitative	16	846	-0.07	-0.07	[-0.24, 0.09]
Male		Biological	Female		Biological	26	3,239	0.15	0.15	[0.15, 0.15]
Male	More	Biological	Female	More	Biological	10	2,636	0.17	0.18	[0.18, 0.18]
Male	Less	Biological	Female	Less	Biological	16	603	0.05	0.05	[0.05, 0.05]
Male		Non-STEM	Female		Non-STEM	26	52,420	0.03	0.03	[-0.12, 0.18]
Male	More	Non-STEM	Female	More	Non-STEM	10	36,049	0.05	0.05	[-0.09, 0.20]
Male	Less	Non-STEM	Female	Less	Non-STEM	16	16,371	-0.02	-0.02	[-0.16, 0.12]

Note: SMC = student major category; *k* = number of institutional studies; *d* = mean observed effect size; δ = estimated mean effect size after corrections for measurement error; CrI = credibility interval.

Table A-4. Estimated Mean HSGPA Differences for SMC Comparisons

Group 1			Group 2							
Gender	Admission Selectivity	SMC	Gender	Admission Selectivity	SMC	*k*	*N*	*d*	*δ*	80% CrI
		Quantitative			Non-STEM	26	58,883	0.42	**0.47**	[0.29, 0.65]
Male		Quantitative	Male		Non-STEM	26	23,846	0.53	**0.60**	[0.40, 0.80]
Female		Quantitative	Female		Non-STEM	26	35,037	0.52	**0.58**	[0.45, 0.72]
	More	Quantitative		More	Non-STEM	10	41,666	0.41	**0.47**	[0.25, 0.68]
	Less	Quantitative		Less	Non-STEM	16	17,217	0.43	**0.48**	[0.40, 0.56]
Male	More	Quantitative	Male	More	Non-STEM	10	17,153	0.52	**0.58**	[0.34, 0.82]
Female	More	Quantitative	Female	More	Non-STEM	10	24,513	0.51	**0.58**	[0.40, 0.76]
Male	Less	Quantitative	Male	Less	Non-STEM	16	6,693	0.57	**0.64**	[0.64, 0.64]
Female	Less	Quantitative	Female	Less	Non-STEM	16	10,524	0.53	**0.60**	[0.60, 0.60]
		Biological			Non-STEM	26	55,659	0.54	**0.61**	[0.45, 0.76]
Male		Biological	Male		Non-STEM	26	20,152	0.64	**0.72**	[0.57, 0.87]
Female		Biological	Female		Non-STEM	26	35,507	0.49	**0.55**	[0.45, 0.64]
	More	Biological		More	Non-STEM	10	36,685	0.52	**0.59**	[0.39, 0.78]
	Less	Biological		Less	Non-STEM	16	16,974	0.58	**0.65**	[0.65, 0.65]
Male	More	Biological	Male	More	Non-STEM	10	13,937	0.60	**0.68**	[0.48, 0.87]
Female	More	Biological	Female	More	Non-STEM	10	24,748	0.48	**0.54**	[0.36, 0.73]
Male	Less	Biological	Male	Less	Non-STEM	16	6,215	0.73	**0.83**	[0.83, 0.83]
Female	Less	Biological	Female	Less	Non-STEM	16	10,759	0.49	**0.55**	[0.55, 0.55]

Note: SMC = student major category; *k* = number of institutional studies; *d* = mean observed effect size; *δ* = estimated mean effect size after corrections for measurement error; CrI = credibility interval; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero.

Table A-5. Estimated Mean HSGPA Differences for Gender Comparisons

Group 1			Group 2							
Gender	Admission Selectivity	SMC	Gender	Admission Selectivity	SMC	*k*	*N*	*d*	*δ*	80% CrI
Male			Female			26	62,122	-0.22	**-0.25**	[-0.38, -0.13]
Male	More		Female	More		10	44,302	-0.19	**-0.21**	[-0.29, -0.14]
Male	Less		Female	Less		16	17,820	-0.31	**-0.35**	[-0.49, -0.21]
Male		Quantitative	Female		Quantitative	26	6,463	-0.29	**-0.33**	[-0.42, -0.23]
Male	More	Quantitative	Female	More	Quantitative	10	5,716	-0.27	**-0.31**	[-0.31, -0.31]
Male	Less	Quantitative	Female	Less	Quantitative	16	846	-0.39	**-0.45**	[-0.76, -0.13]
Male		Biological	Female		Biological	26	3,239	-0.13	-0.14	[-0.27, -0.01]
Male	More	Biological	Female	More	Biological	10	2,636	-0.13	-0.15	[-0.31, 0.02]
Male	Less	Biological	Female	Less	Biological	16	603	-0.11	-0.12	[-0.12, -0.12]
Male		Non-STEM	Female		Non-STEM	26	52,420	-0.31	**-0.35**	[-0.46, -0.24]
Male	More	Non-STEM	Female	More	Non-STEM	10	36,049	-0.29	**-0.33**	[-0.41, -0.24]
Male	Less	Non-STEM	Female	Less	Non-STEM	16	16,371	-0.36	**-0.40**	[-0.53, -0.28]

Note: SMC = student major category; *k* = number of institutional studies; *d* = mean observed effect size; *δ* = estimated mean effect size after corrections for measurement error; CrI = credibility interval; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero.

Table A-6. Estimated Mean Semester GPA Differences for STEM-Quantitative versus Non-STEM Comparisons

Admission Selectivity	*k*	Semester	Male and Female			Male			Female		
			d	δ	80% CrI	*d*	δ	80% CrI	*d*	δ	80% CrI
All Levels	26	1	0.33	**0.43**	[0.21, 0.66]	0.44	**0.58**	[0.32, 0.84]	0.42	**0.54**	[0.54, 0.54]
		2	0.16	0.21	[-0.06, 0.48]	0.27	**0.36**	[0.09, 0.64]	0.22	**0.28**	[0.05, 0.51]
		3	0.02	0.03	[-0.28, 0.34]	0.13	0.18	[-0.16, 0.51]	0.09	0.12	[-0.03, 0.26]
		4	-0.05	-0.06	[-0.45, 0.32]	0.06	0.09	[-0.31, 0.49]	0.00	0.00	[-0.32, 0.31]
		5	-0.11	-0.14	[-0.50, 0.23]	0.02	0.03	[-0.36, 0.42]	-0.07	-0.08	[-0.36, 0.19]
		6	-0.09	-0.11	[-0.40, 0.18]	0.03	0.05	[-0.24, 0.34]	-0.02	-0.02	[-0.27, 0.23]
		7	-0.11	-0.14	[-0.39, 0.10]	0.04	0.05	[-0.20, 0.30]	-0.07	-0.09	[-0.36, 0.18]
		8	-0.08	-0.10	[-0.35, 0.14]	0.07	0.08	[-0.18, 0.35]	-0.03	-0.04	[-0.04, -0.04]
More-Selective	10	1	0.35	**0.45**	[0.37, 0.54]	0.44	**0.58**	[0.45, 0.71]	0.44	**0.57**	[0.57, 0.57]
		2	0.17	**0.22**	[0.06, 0.38]	0.27	**0.36**	[0.17, 0.54]	0.22	**0.28**	[0.10, 0.46]
		3	0.00	0.01	[-0.27, 0.29]	0.11	0.16	[-0.16, 0.48]	0.04	0.05	[-0.06, 0.15]
		4	-0.10	-0.12	[-0.51, 0.27]	0.02	0.04	[-0.37, 0.44]	-0.07	-0.09	[-0.39, 0.20]
		5	-0.14	-0.18	[-0.55, 0.20]	0.00	0.01	[-0.37, 0.39]	-0.11	-0.14	[-0.42, 0.15]
		6	-0.09	-0.12	[-0.41, 0.17]	0.02	0.04	[-0.27, 0.34]	-0.04	-0.04	[-0.23, 0.15]
		7	-0.11	-0.14	[-0.36, 0.08]	0.04	0.04	[-0.19, 0.28]	-0.05	-0.07	[-0.16, 0.03]
		8	-0.08	-0.11	[-0.35, 0.14]	0.07	0.08	[-0.19, 0.35]	-0.04	-0.06	[-0.19, 0.08]
Less-Selective	16	1	0.30	**0.39**	[0.00*, 0.78]	0.43	**0.57**	[0.12, 1.02]	0.37	**0.47**	[0.47, 0.47]
		2	0.15	0.19	[-0.24, 0.63]	0.28	0.38	[-0.04, 0.80]	0.22	0.29	[-0.03, 0.61]
		3	0.07	0.09	[-0.28, 0.46]	0.18	0.23	[-0.14, 0.60]	0.21	**0.27**	[0.27, 0.27]
		4	0.06	0.08	[-0.23, 0.39]	0.18	0.24	[-0.09, 0.56]	0.16	**0.20**	[0.02, 0.39]
		5	-0.04	-0.05	[-0.36, 0.26]	0.08	0.09	[-0.30, 0.48]	0.03	0.03	[-0.16, 0.22]
		6	-0.07	-0.09	[-0.38, 0.20]	0.06	0.08	[-0.17, 0.32]	0.02	0.03	[-0.32, 0.37]
		7	-0.12	-0.15	[-0.44, 0.14]	0.05	0.06	[-0.22, 0.34]	-0.11	-0.14	[-0.60, 0.32]
		8	-0.07	-0.09	[-0.33, 0.15]	0.07	0.09	[-0.17, 0.34]	0.02	0.02	[0.02, 0.02]

Note: *k* = number of institutional studies; *d* = mean observed effect size; δ = estimated mean effect size after corrections for measurement error; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero; *lower limit = 0.002, rounded to 0.00.

Table A-7. Estimated Mean Semester GPA Differences for STEM-Biological versus Non-STEM Comparisons

Admission Selectivity	*k*	Semester	Male and Female *d*	Male and Female *δ*	Male and Female 80% CrI	Male *d*	Male *δ*	Male 80% CrI	Female *d*	Female *δ*	Female 80% CrI
All Levels	26	1	0.46	**0.59**	[0.35, 0.84]	0.53	**0.70**	[0.34, 1.06]	0.42	**0.54**	[0.40, 0.69]
		2	0.39	**0.51**	[0.25, 0.76]	0.50	**0.66**	[0.44, 0.88]	0.33	**0.43**	[0.23, 0.63]
		3	0.33	**0.42**	[0.20, 0.64]	0.49	**0.63**	[0.46, 0.80]	0.25	**0.31**	[0.11, 0.51]
		4	0.27	**0.34**	[0.18, 0.51]	0.41	**0.53**	[0.38, 0.68]	0.19	**0.24**	[0.16, 0.31]
		5	0.20	**0.26**	[0.10, 0.42]	0.41	**0.52**	[0.41, 0.63]	0.08	0.11	[-0.01, 0.22]
		6	0.17	**0.21**	[0.04, 0.38]	0.40	**0.50**	[0.35, 0.66]	0.04	0.05	[-0.15, 0.25]
		7	0.08	0.09	[-0.10, 0.29]	0.27	**0.34**	[0.02, 0.66]	-0.04	-0.05	[-0.17, 0.08]
		8	0.05	0.06	[-0.17, 0.29]	0.18	0.23	[-0.08, 0.54]	-0.03	-0.04	[-0.31, 0.24]
More-Selective	10	1	0.50	**0.64**	[0.45, 0.84]	0.62	**0.81**	[0.56, 1.07]	0.44	**0.57**	[0.43, 0.70]
		2	0.43	**0.56**	[0.29, 0.83]	0.53	**0.70**	[0.42, 0.98]	0.38	**0.48**	[0.25, 0.72]
		3	0.37	**0.47**	[0.28, 0.67]	0.51	**0.65**	[0.47, 0.84]	0.30	**0.37**	[0.23, 0.51]
		4	0.30	**0.38**	[0.18, 0.58]	0.45	**0.57**	[0.34, 0.81]	0.21	**0.27**	[0.17, 0.38]
		5	0.24	**0.31**	[0.13, 0.49]	0.42	**0.54**	[0.33, 0.75]	0.14	0.18	[0.08, 0.27]
		6	0.21	**0.27**	[0.11, 0.43]	0.39	**0.50**	[0.41, 0.59]	0.11	0.14	[0.00, 0.29]
		7	0.07	0.09	[-0.10, 0.28]	0.23	0.29	[-0.02, 0.59]	-0.03	-0.03	[-0.03, -0.03]
		8	0.05	0.07	[-0.19, 0.32]	0.18	0.23	[-0.07, 0.53]	-0.02	-0.03	[-0.22, 0.17]
Less-Selective	16	1	0.37	**0.48**	[0.19, 0.76]	0.34	**0.45**	[0.10, 0.81]	0.38	**0.49**	[0.35, 0.64]
		2	0.30	**0.39**	[0.32, 0.46]	0.43	**0.57**	[0.57, 0.57]	0.23	**0.29**	[0.29, 0.29]
		3	0.24	**0.31**	[0.11, 0.52]	0.44	**0.57**	[0.47, 0.66]	0.13	0.17	[-0.03, 0.37]
		4	0.20	**0.25**	[0.25, 0.25]	0.33	**0.43**	[0.43, 0.43]	0.12	0.15	[0.15, 0.15]
		5	0.12	0.14	[0.14, 0.14]	0.38	**0.48**	[0.48, 0.48]	-0.04	-0.05	[-0.05, -0.05]
		6	0.07	0.09	[0.09, 0.09]	0.41	**0.51**	[0.28, 0.75]	-0.12	-0.15	[-0.15, -0.15]
		7	0.09	0.11	[-0.10, 0.32]	0.38	**0.47**	[0.17, 0.76]	-0.06	-0.08	[-0.30, 0.15]
		8	0.04	0.04	[-0.11, 0.20]	0.19	0.23	[-0.10, 0.56]	-0.05	-0.06	[-0.24, 0.11]

Note: k = number of institutional studies; *d* = mean observed effect size; *δ* = estimated mean effect size after corrections for measurement error; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero.

Table A-8. Estimated Mean Semester GPA Differences for Male versus Female Comparisons, Overall and by Admission Selectivity

Admission Selectivity	*k*	Semester	*d*	*δ*	80% CrI
All Levels	26	1	-0.19	**-0.24**	[-0.38, -0.11]
		2	-0.21	**-0.27**	[-0.40, -0.15]
		3	-0.22	**-0.28**	[-0.39, -0.18]
		4	-0.22	**-0.29**	[-0.38, -0.20]
		5	-0.27	**-0.35**	[-0.44, -0.25]
		6	-0.25	**-0.32**	[-0.43, -0.21]
		7	-0.31	**-0.38**	[-0.46, -0.30]
		8	-0.29	**-0.36**	[-0.45, -0.28]
More-Selective	10	1	-0.14	-0.19	[-0.28, -0.10]
		2	-0.17	**-0.22**	[-0.31, -0.14]
		3	-0.19	**-0.24**	[-0.32, -0.16]
		4	-0.20	**-0.26**	[-0.32, -0.20]
		5	-0.26	**-0.33**	[-0.42, -0.24]
		6	-0.23	**-0.30**	[-0.40, -0.19]
		7	-0.29	**-0.36**	[-0.43, -0.29]
		8	-0.28	**-0.34**	[-0.43, -0.26]
Less-Selective	16	1	-0.29	**-0.38**	[-0.41, -0.35]
		2	-0.30	**-0.39**	[-0.43, -0.35]
		3	-0.30	**-0.38**	[-0.43, -0.33]
		4	-0.28	**-0.36**	[-0.43, -0.29]
		5	-0.31	**-0.39**	[-0.45, -0.32]
		6	-0.30	**-0.38**	[-0.49, -0.27]
		7	-0.35	**-0.44**	[-0.50, -0.37]
		8	-0.33	**-0.41**	[-0.46, -0.36]

Note: *k* = number of institutional studies; *d* = mean observed effect size; *δ* = estimated mean effect size after corrections for measurement error; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero.

Table A-9. Estimated Mean Semester GPA Differences for Male versus Female Student Comparisons, by SMC and Admission Selectivity

			STEM-Quantitative			STEM-Biological			Non-STEM		
Admission Selectivity	*k*	Semester	*d*	δ	80% CrI	*d*	δ	80% CrI	*d*	δ	80% CrI
All Levels	26	1	-0.24	**-0.31**	[-0.31, -0.31]	-0.10	-0.13	[-0.35, 0.10]	-0.26	**-0.33**	[-0.43, -0.24]
		2	-0.17	**-0.23**	[-0.23, -0.23]	-0.07	-0.08	[-0.08, -0.08]	-0.26	**-0.33**	[-0.45, -0.22]
		3	-0.17	**-0.21**	[-0.34, -0.08]	0.00	-0.01	[-0.07, 0.04]	-0.25	**-0.32**	[-0.41, -0.23]
		4	-0.15	-0.18	[-0.26, -0.10]	0.00	0.00	[0.00, 0.00]	-0.24	**-0.31**	[-0.39, -0.23]
		5	-0.18	**-0.24**	[-0.30, -0.17]	0.02	0.02	[0.02, 0.02]	-0.29	**-0.37**	[-0.46, -0.28]
		6	-0.18	**-0.22**	[-0.39, -0.06]	0.05	0.06	[-0.16, 0.28]	-0.27	**-0.35**	[-0.45, -0.24]
		7	-0.20	**-0.25**	[-0.30, -0.20]	-0.04	-0.05	[-0.25, 0.14]	-0.33	**-0.41**	[-0.50, -0.33]
		8	-0.20	**-0.24**	[-0.24, -0.24]	-0.10	-0.12	[-0.25, 0.00]	-0.32	**-0.40**	[-0.47, -0.32]
More	10	1	-0.23	**-0.30**	[-0.30, -0.30]	-0.05	-0.06	[-0.20, 0.09]	-0.23	**-0.30**	[-0.39, -0.21]
		2	-0.17	**-0.22**	[-0.22, -0.22]	-0.06	-0.07	[-0.07, -0.07]	-0.23	**-0.30**	[-0.40, -0.19]
		3	-0.14	-0.17	[-0.26, -0.09]	0.00	-0.01	[-0.01, -0.01]	-0.22	**-0.29**	[-0.35, -0.22]
		4	-0.13	-0.16	[-0.22, -0.10]	0.02	0.02	[0.02, 0.02]	-0.22	**-0.28**	[-0.33, -0.23]
		5	-0.17	**-0.22**	[-0.22, -0.22]	0.01	0.01	[0.01, 0.01]	-0.28	**-0.36**	[-0.45, -0.26]
		6	-0.17	**-0.21**	[-0.38, -0.05]	0.03	0.03	[0.03, 0.03]	-0.25	**-0.32**	[-0.42, -0.23]
		7	-0.20	**-0.25**	[-0.28, -0.23]	-0.07	-0.09	[-0.26, 0.08]	-0.32	**-0.39**	[-0.46, -0.32]
		8	-0.19	**-0.23**	[-0.29, -0.18]	-0.09	-0.12	[-0.12, -0.12]	-0.31	**-0.39**	[-0.46, -0.32]
Less	16	1	-0.30	**-0.40**	[-0.40, -0.40]	-0.35	**-0.44**	[-0.44, -0.44]	-0.31	**-0.41**	[-0.41, -0.41]
		2	-0.22	**-0.30**	[-0.30, -0.30]	-0.09	-0.11	[-0.11, -0.11]	-0.32	**-0.41**	[-0.47, -0.36]
		3	-0.35	**-0.44**	[-0.44, -0.44]	-0.01	-0.04	[-0.38, 0.29]	-0.31	**-0.40**	[-0.44, -0.35]
		4	-0.26	**-0.33**	[-0.33, -0.33]	-0.07	-0.10	[-0.21, 0.01]	-0.29	**-0.38**	[-0.45, -0.31]
		5	-0.25	**-0.35**	[-0.58, -0.12]	0.07	0.08	[-0.05, 0.21]	-0.32	**-0.40**	[-0.46, -0.35]
		6	-0.21	**-0.29**	[-0.45, -0.13]	0.14	0.15	[-0.34, 0.65]	-0.31	**-0.39**	[-0.51, -0.28]
		7	-0.20	**-0.25**	[-0.38, -0.12]	0.07	0.09	[-0.08, 0.26]	-0.37	**-0.46**	[-0.55, -0.37]
		8	-0.23	**-0.28**	[-0.28, -0.28]	-0.14	-0.15	[-0.52, 0.22]	-0.34	**-0.42**	[-0.48, -0.36]

Note: *k* = number of institutional studies; *d* = mean observed effect size; δ = estimated mean effect size after corrections for measurement error; **bold** = effect size exceeds |0.20| and the credibility interval does not contain zero.

Table A-10. Estimated Mean Differences for SMC Comparisons: ACT Composite Score, HSGPA, First-Semester GPA

Group 1			Group 2			δ			
Gender	Admission Selectivity	SMC	Gender	Admission Selectivity	SMC	ACT Composite	HSGPA	First-Semester GPA	Eighth-Semester GPA
		Quantitative			Non-STEM	**0.77**	**0.47**	**0.43**	-0.10
Male		Quantitative	Male		Non-STEM	**0.74**	**0.60**	**0.58**	0.08
Female		Quantitative	Female		Non-STEM	**0.79**	**0.58**	**0.54**	-0.04
	More	Quantitative		More	Non-STEM	**0.81**	**0.47**	**0.45**	-0.11
	Less	Quantitative		Less	Non-STEM	**0.68**	**0.48**	**0.39**	-0.09
Male	More	Quantitative	Male	More	Non-STEM	**0.75**	**0.58**	**0.58**	0.08
Female	More	Quantitative	Female	More	Non-STEM	**0.85**	**0.58**	**0.57**	-0.06
Male	Less	Quantitative	Male	Less	Non-STEM	**0.72**	**0.64**	**0.57**	0.09
Female	Less	Quantitative	Female	Less	Non-STEM	**0.66**	**0.60**	**0.47**	0.02
		Biological			Non-STEM	**0.66**	**0.61**	**0.59**	0.06
Male		Biological	Male		Non-STEM	**0.73**	**0.72**	**0.70**	0.23
Female		Biological	Female		Non-STEM	**0.61**	**0.55**	**0.54**	-0.04
	More	Biological		More	Non-STEM	**0.68**	**0.59**	**0.64**	0.07
	Less	Biological		Less	Non-STEM	**0.60**	**0.65**	**0.48**	0.04
Male	More	Biological	Male	More	Non-STEM	**0.76**	**0.68**	**0.81**	0.23
Female	More	Biological	Female	More	Non-STEM	**0.64**	**0.54**	**0.57**	-0.03
Male	Less	Biological	Male	Less	Non-STEM	**0.67**	**0.83**	**0.45**	0.23
Female	Less	Biological	Female	Less	Non-STEM	**0.56**	**0.55**	**0.49**	-0.06

Note: SMC = student major category; δ = estimated mean effect size after corrections for measurement error; **bold** = effect size exceeds |0.20| and the credibility interval did not contain zero.

Table A-11. Estimated Mean Differences for Gender Comparisons: ACT Composite Score, HSGPA, Semester GPA (average)

Group 1			Group 2			δ		
Gender	Admission Selectivity	SMC	Gender	Admission Selectivity	SMC	ACT Composite	HSGPA	Semester GPA (average)
Male			Female			0.14	**-0.25**	**-0.31**
Male	More		Female	More		0.18	**-0.21**	**-0.28**
Male	Less		Female	Less		0.03	**-0.35**	**-0.39**
Male		Quantitative	Female		Quantitative	-0.02	**-0.33**	**-0.24**
Male	More	Quantitative	Female	More	Quantitative	-0.01	**-0.31**	**-0.22**
Male	Less	Quantitative	Female	Less	Quantitative	-0.07	**-0.45**	**-0.33**
Male		Biological	Female		Biological	0.15	-0.14	-0.04
Male	More	Biological	Female	More	Biological	0.18	-0.15	-0.03
Male	Less	Biological	Female	Less	Biological	0.05	-0.12	-0.06
Male		Non-STEM	Female		Non-STEM	0.03	**-0.35**	**-0.35**
Male	More	Non-STEM	Female	More	Non-STEM	0.05	**-0.33**	**-0.33**
Male	Less	Non-STEM	Female	Less	Non-STEM	-0.02	**-0.40**	**-0.41**

Note: SMC = student major category; δ = estimated mean effect size after corrections for measurement error; **bold** = average effect size exceeds |0.20| and none of the eight credibility intervals contain zero.

ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

For more information, visit www.act.org.

050208150 Rev 1