

Article

Analysis of Student Performance in Large-Enrollment Life Science Courses

Leah Renée Creech and Ryan D. Sweeder

Lyman Briggs College, Michigan State University, East Lansing, MI 48825

Submitted February 2, 2012; Revised August 4, 2012; Accepted August 6, 2012

Monitoring Editor: Jose Herrera

This study examined the historical performance of students at Michigan State University in 12 life sciences courses over 13 yr to find variables impacting student success. Hierarchical linear modeling predicted 25.0–62.8% of the variance in students' grades in the courses analyzed. The primary predictor of a student's course grade was his or her entering grade point average; except for the second course in a series (i.e., Biochemistry II), in which the grade for the first course in the series (i.e., Biochemistry I) was often the best predictor, as judged by β values. Student gender and major were also statistically significant for a majority of the courses studied. Female students averaged grades 0.067–0.303 lower than their equivalent male counterparts, and majors averaged grades were 0.088–0.397 higher than nonmajors. Grades earned in prerequisite courses provided minimal predictive ability. Ethnicity and involvements in honors college or science residential college were generally insignificant.

INTRODUCTION

There have been many calls to improve the quality of undergraduate science education. *BIO2010* (National Research Council, 2003) and the Boyer Commission (1998) focused on the enhancement of research opportunities at universities in order to improve science education, increase diverse experiences, and prepare the next generation of scientific researchers. *Rising above the Gathering Storm* (National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine, 2007) suggested that similar actions be taken in higher education and research to restore the scientific and technological foundation of the U.S. economy. Answering these calls requires a better understanding of approaches that lead to student success and enhanced learning.

A number of studies have focused on understanding the factors that lead to student success and subsequently enhance student learning (Cheng and Ickes, 2009; Miyake *et al.*, 2010).

Elaine Seymour postulated that students internalize experiences about their success in science, and this internalization starts early in a student's academic career and builds up to create a perception that often negatively affects a student's persistence in the sciences (Seymour and Hewitt, 1997). This phenomenon is known as the "leaky pipeline." The number of students who pursue sciences throughout their undergraduate education has been found to decrease by 40–60% (Seymour and Hewitt, 1997), with life sciences having a net loss of nearly 50% (Astin, 1993). Many students have difficulty transitioning from introductory and advanced science courses and are met with negative experiences. Science faculty at many universities have been known to teach in a more authoritarian style, with little encouragement of classroom discussion; to use only multiple-choice exams; and to exhibit less interest in students' personal development (Astin, 1993). Students who meet with these negative experiences throughout their undergraduate years are more likely to feel they are unable to compete in the sciences and often switch to a different field (Griffith, 2010).

Recent studies indicate that students receive a variety of these negative signals in introductory courses. Variation in the grades earned in introductory science courses is one factor that affects students' persistence in science-related majors; a lightly graded introductory course versus a harshly graded one has a different effect (Ost, 2010). Other factors include performance in prerequisite courses (Seymour and Hewitt, 1997; Turner and Lindsay, 2003), overall academic achievement, ethnicity (Seymour and Hewitt, 1997), and other

DOI: 10.1187/cbe.12-02-0019

Address correspondence to: Ryan D. Sweeder (sweeder@msu.edu).

© 2012 L. R. Creech and R. D. Sweeder. CBE—Life Sciences Education © 2012 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

experiences, including participation in residential and honors colleges (Astin, 1993). Factors such as motivation, conscientiousness, or confidence (Brownlow *et al.*, 2000) can also impact student performance, but these are more challenging to measure. Gender is another factor in the leaky pipeline (Seymour and Hewitt, 1997), since male and female students are impacted differently in similar situations (Fox and Firebaugh, 1992; Beyer, 1999). One might imagine that students who are persistent and academically successful enough to navigate introductory science courses might be expected to meet with equivalent success in upper-level courses, regardless of gender, major, ethnic, and background differences. Nonetheless, not all of the factors listed would be expected to have equal significance. Despite compelling research findings in specific courses, little has been done to systematically compare male and female students' performance in large-enrollment, upper-level science courses across an *entire* university. Although students may ignore one undesirable grade, repeated low grades are likely to encourage them to leave the sciences due to lack of confidence in their own abilities and/or a decrease in motivation (Astin, 1993).

We have previously reported a performance difference in a single biochemistry course at Michigan State University (MSU; Rauschenberger and Sweeder, 2010). Using hierarchical linear modeling (HLM), we found that the most important factor was cumulative grade point average (GPA), although student gender and major also were significant. This study is an extension of the biochemistry research using HLM to determine the best predictors of student performance in many life sciences courses.

METHODS

With the approval of the MSU Institutional Review Board (IRB #07-446), student data were collected from the 21,688 students who completed upper-level science courses at Michigan State University from 1997 to 2010. The science courses used were Physiology I (PHY 431) and II (PHY 432), Basic Biochemistry (BC 401), Biochemistry I (BC 461) and II (BC 462), Introductory Microbiology (MB 201), Advanced Introductory Microbiology (MB 301), Genetics (Gen 341), Organic Chemistry I (Orgo 251) and II (Orgo 252), and Advanced Organic Chemistry I (Orgo 351) and II (Orgo 352). The data set consisted of grades in upper-level science courses, introductory courses (biology, chemistry, and organic chemistry), entering GPA by semester, major, honors college or Lyman Briggs College (LBC), and demographic information (gender and ethnicity). Students were considered LBC students if they were enrolled in any of the introductory LBC courses.

Student performance in life sciences courses was measured by the grade achieved during the student's first enrollment in each course. HLM analyses (using SPSS 19; SPSS, 2010) created mathematical models to predict student performance in different life sciences courses based on specific elements. Different combinations of independent variables (italicized throughout) used as modeling predictors were *Entering GPA* for the course, *Course Grade for General Chemistry I and II, Biology I and II, Organic I and II*, and the first of a series of courses (e.g., PHY 431 performance was used as a predictor for performance in PHY 432). Categorical variables included *Gender* (Seymour and Hewitt, 1997), *Ethnicity* (Seymour and

Hewitt, 1997), *Major* (Ost, 2010; Solnick, 1995), enrollment in *LBC* (Inkelas *et al.*, 2008; Stassen, 2003) or *Honors College*, and enrollment in specific prerequisite science courses. Categorical variables were used to split all students into one of two groups using dummy variables (e.g., female and nonfemale for the *Gender* variable). All models included *Entering GPA* as a predictor, and most models included all of the listed categorical variables.

HLM created a model for predicting students' life sciences grades by fitting the following equation for each student:

$$\text{Grade} = b_1x_1 + b_2x_2 + \dots b_ix_i + \varepsilon$$

where the b_i values are the coefficients determined by the model, the x_i values are the factors entered into the model, and ε represents the residual error in the prediction for the specific student. This method fits the coefficients to minimize total error, which leads to the amount of variance predicted, R^2 . The advantage of the model is that it will statistically account for incoming differences in student abilities among groups. Because coefficients scale inversely with the range of their factors, direct examination of the values are created by subtracting the sample mean from a score and subsequently dividing it by the SD. These new values are then used in a regression analysis to determine the β coefficient, which allows us to compare the relative significance of a variable in contributing to the model by measuring all variables on an equivalent scale. β coefficients range from -1 to $+1$, with the sign indicating a positive or negative relationship. In the stepwise analysis, t tests are used to indicate whether a variable has a significant impact independent of the other variables already present in the model. Variables are added in order of greatest significance, until no additional factors have a significant influence on the predictive power of the model (Rauschenberger and Sweeder, 2010). HLM used our raw data and was executed stepwise, excluding cases list-wise; therefore, not all students who were enrolled in a particular course were used in all analyses.

With HLM, factors known to affect the prediction of student performance are used to create a model. A second set of factors are then added to determine if they provide any additional explanation of variance. Graphical representations of the data were created through SPSS 19 to highlight categorical differences in the data. Average grades for different student groups in a specific course are compared as clusters and are based on *Entering GPA* using scatter plots. Students' GPAs were clustered to the nearest quarter grade in the plots, and the error bars represent a 95% confidence level.

RESULTS AND DISCUSSION

Entering GPA tended to be the variable most highly correlated to student grades in each course studied (Table 1), similar to previous results (Wright *et al.*, 2009; Rauschenberger and Sweeder, 2010). When each class was analyzed using our HLM, *Entering GPA* was the single best predictor, accounting for the most variance in all models (22.3–58.0%), followed frequently by *Gender* (1–3%). *Entering GPA* was used in the HLM, because of its strong predictive ability in performance; it represents many important characteristics that demonstrate student success, such as academic aptitude, study skills, background, and motivation. For the most basic of our models,

Table 1. Partial summary of HLM for 12 life sciences courses^a

| | Physiology | | | Biochemistry | | | Molecular biology | | | Organic chemistry | | | Genetics |
|---------------------------------|------------|-----------|--|--------------|--------|-------|-------------------|--------|--------|-------------------|--------|-------|----------|
| | 431 | 432 | | 401 | 461 | 462 | 201 | 301 | 251 | 252 | 351 | 352 | 341 |
| R^2 | 0.390 | 0.628 | | 0.316 | 0.465 | 0.600 | 0.429 | 0.412 | 0.368 | 0.516 | 0.250 | 0.522 | 0.381 |
| N | 5728 | 5185 | | 4161 | 6381 | 5768 | 1466 | 4890 | 13,494 | 11,313 | 2068 | 2083 | 7337 |
| (approximate class size) | (350–500) | (300–500) | | (150) | (350) | (350) | (300) | (200) | (350) | (350) | (200) | (175) | (250) |
| β values for: | | | | | | | | | | | | | |
| <i>Entering GPA</i> | 0.565 | 0.315 | | 0.568 | 0.662 | 0.336 | 0.636 | 0.623 | 0.538 | 0.381 | 0.428 | 0.387 | 0.580 |
| First series of course | N/A | 0.528 | | N/A | N/A | 0.491 | N/A | N/A | N/A | 0.389 | N/A | 0.392 | N/A |
| <i>Female</i> | –0.133 | –0.032 | | –0.053 | –0.067 | INSIG | INSIG | –0.074 | –0.092 | –0.026 | –0.061 | INSIG | –0.021 |
| <i>Major</i> | 0.114 | 0.036 | | N/A | 0.072 | 0.027 | N/A | 0.085 | N/A | N/A | N/A | N/A | N/A |
| Unstandardized coefficient for: | | | | | | | | | | | | | |
| <i>Female</i> | –0.303 | –0.072 | | –0.117 | –0.135 | INSIG | INSIG | –0.165 | –0.224 | –0.067 | –0.139 | INSIG | –0.044 |
| <i>Major</i> | 0.289 | 0.088 | | N/A | 0.397 | 0.130 | N/A | 0.293 | N/A | N/A | N/A | N/A | N/A |

^a Adjusted R^2 represents the decimal percent of variance explained. *Major* represents those students who are majoring in that subject (e.g., biochemistry majors for BC 461 or physiology majors for PHY 431). Both *Gender* and *Major* are dummy variables, so the unstandardized coefficient represents the difference between being in the category or not. N/A = Not applicable. INSIG = Not significant.

all students who took a given class were included. However, when examining the impact of introductory courses, students who met the prerequisites through Advanced Placement exams or transfer credit were not included, and the results should not be applied to this student subpopulation.

Ethnicity did not show any clear impact on performance prediction in any HLM. Occasionally, certain ethnic groups, such as African Americans or Asians, had a significant impact on performance prediction in a specific class, but no trends in one ethnic group emerged in the modeling. Similarly, *High school GPA* and *ACT/SAT scores* also did not show any significance in the HLM once GPA was accounted for and were removed from the subsequent models.

Major Factors

The best predictive factor in this study was, not surprisingly, the student's cumulative GPA (*Entering GPA*) upon entering the course under analysis. *Gender* was the second most important factor for all courses examined, except BC 462, MB 201, Gen 341, and Orgo 352, which showed no significant differences between genders. The β value of *Gender* ranged up to one-quarter of the value of *Entering GPA*, indicating it is up to one-quarter as important (Table 1). BC 462 and Orgo 352 may have showed no gender differences because *BC 461 grade* and *Orgo 351 grade*, respectively, were used as predictors; these grades already include the gender differences from the first course in the series. However, female students continued to display lower performance in PHY 432 and Orgo 252, even though the *PHY 431 grade* or *Orgo 251 grade* was used as a predictor.

When course performance was analyzed using HLM, PHY 431 had the greatest gender performance differences, followed by Orgo 251 (Table 1). A graph of average GPA in PHY 431 versus students' *Entering GPA* provides visual representation to better demonstrate the impact of *Gender* (Figure 1a). This graph highlights that the gender performance differences span the entering GPA spectrum and are not just present in either high- or low-performing students. By contrast, a similar graph of MB 201, a class with insignificant gender differences, shows both genders performing equivalently over the span of *Entering GPAs* (Figure 1b).

Role of Student Major

Students in classes closely linked to their major may be expected to have higher levels of motivation and interest in the subject, which may result in higher performance. To evaluate this possibility, we included a variable to identify students based on their college major (Table 1). *Major* resulted in a β value that was up to one-fifth the value of *Entering GPA*, slightly less than the impact of gender. The unstandardized coefficients in the models indicate that being a major in the field of a course yields a positive boost that is on par with or larger than the negative adjustment for being female (Table 1).

The impacts of gender and major may be linked due to uneven distribution of male and female students across the majors. For testing this hypothesis, human biology, physiology, biochemistry, microbiology, and zoology majors were individually analyzed, yet typically still showed the gender gap (Table 2). Human biology and physiology majors exhibited the gender gap in certain classes (PHY 431, BC 461, Orgo

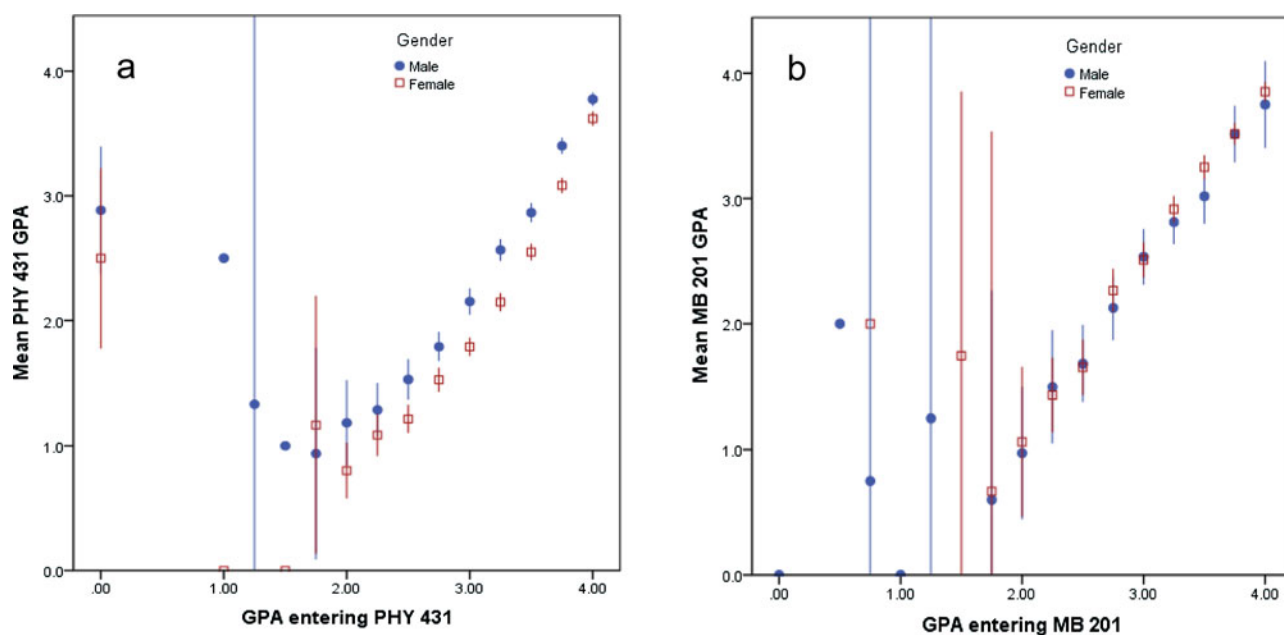


Figure 1. Gender performance differences in PHY 431 vs. MB 201: (a) Average course GPA in PHY 431 compared with average *Entering GPA* for PHY 431, separated by *Gender*. (b) Average course GPA in MB 201 compared with average *Entering GPA* for MB 201, separated by *Gender*. In both models, blue represents male students and red represents female students.

251), while biochemistry majors showed no gender gap in any of the courses analyzed (Table 2).

Impact of Honors College, Residential College, Introductory Courses, and Math Readiness

It is possible that highly motivated students may exhibit different trends than the population as a whole. Students in the honors college or the residential college may fit this category, yet HLM analysis yielded little evidence of performance differences. Similarly, we examined the performance differences in introductory classes and found that they had relatively little impact on performance in upper-level courses. This result is similar to the work of Wright *et al.* (2009), which indicated

that an organic chemistry prerequisite had no impact on a biochemistry grade. Mathematics and physics courses were not analyzed in our model, and it is possible that gender was acting as a proxy for this variable, as females have been found to not perform as well in prerequisite math courses (Brownlow *et al.*, 2000). However, when ACT/SAT subscores were included as a surrogate for student math readiness, both showed up as insignificant in modeling when *Entering GPA* was used, which led us to reject this hypothesis.

Historical Trends of Gender Differences

We examined the gender gap in each course between 1997 and 2009. Although the gap in most classes seemed to be

Table 2. Female performance compared with males within life sciences majors^a

| Course | Physiology | Human biology | Microbiology | Biochemistry | Zoology |
|-----------------------|------------|---------------|--------------|--------------|---------|
| Physiology 431 | −0.311 | −0.289 | | | |
| Physiology 432 | | −0.069 | | | |
| Biochemistry 401 | | | | N/A | |
| Biochemistry 461 | −0.146 | −0.116 | −0.145 | | −0.335 |
| Biochemistry 462 | | | | | 0.203 |
| Molecular Biology 201 | | | N/A | N/A | N/A |
| Molecular Biology 301 | | −0.113 | −0.162 | | −0.255 |
| Organic 251 | −0.285 | −0.185 | −0.187 | | −0.155 |
| Organic 252 | | | | | |
| Organic 351 | | | | | N/A |
| Organic 352 | | | | | |
| Genetics 341 | | | | | −0.088 |

^aHLM coefficients representing female performance relative to male students within a single major (column headings). Blank spaces indicate that *Gender* was an insignificant variable. N/A: students in the major are not required and typically do not take the course listed; therefore no HLM could be produced. Red coefficients indicate HLMs for which *Gender* explained 1–3% of the variance; black coefficients indicate that <1% of variance was added.

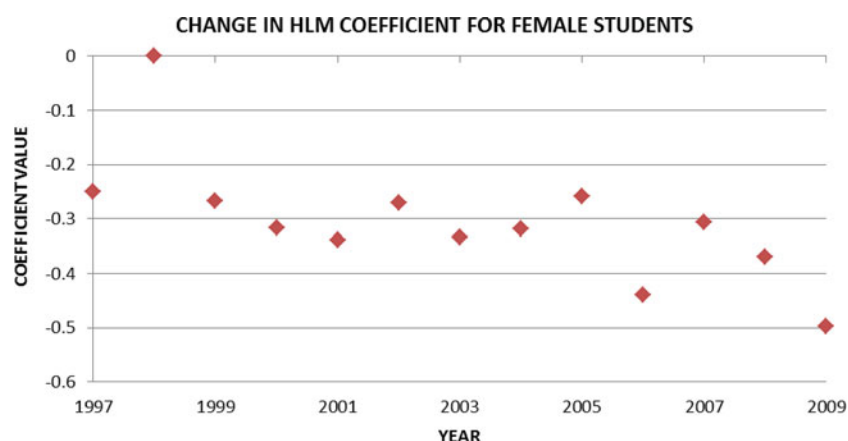


Figure 2. Historical trend of deteriorating female performance in PHY 431. The HLM coefficient represents the model's performance gap between male and female students. No significant difference was found in 1998.

relatively consistent during this time, similar to previous work (Nowell and Hedges, 1998), PHY 431 was an interesting exception (Figure 2). The calculated negative coefficient for females has increased over the past decade, reaching a difference in grades of 0.5 between male and female students in 2009. Other courses with a significant gender gap in the HLM showed consistency in the gender gap over the past decade. One possible explanation is that the enrollment in PHY 431 has increased during this time span from around 300 students (1997–2001) to more than 525 students (2005–2009). The greater difference may simply reflect that females are more negatively impacted by large class size (Seymour and Hewitt, 1997).

IMPLICATIONS OF RESULTS

The findings that better students perform better in life sciences courses or that majors perform better in classes in their discipline is not surprising. However, it is surprising that females were found to perform at a lower level in the classes studied, given their better overall cumulative GPA. Simply comparing the average score or grade in the course cannot accurately assess whether the two genders are performing equivalently. Instead, a more robust statistical approach, such as HLM, is required to definitively identify gender performance differences. Unlike most studies that highlight bias within a single class, this study illustrates such disparity across the curriculum, with male students outperforming equivalently prepared females (as measured by *Entering GPA* and previous science course grades). This trend is worrying, as we consider the signals that we are sending to our highly capable female students working toward a career in the sciences. In a wide array of courses throughout their academic career, they are earning lower grades compared with their male counterparts and quietly receiving signals from this that they are not as successful in the field. Given the results of this study, how should we think about addressing this aspect of the leaky scientific pipeline?

The first step to address the leaky pipeline is to identify the sources. There have been many factors known to lead to the potential gender gap in science courses, including: how the class is formatted (Astin, 1993), the professor's gender (Seymour and Hewitt, 1997), the size of the class (Kokkelenberg *et al.*, 2008), and the format of the assessment

(Seymour and Hewitt, 1997). It is interesting to note that whatever the reason for the bias in these classes, it cuts across the range of students' abilities and majors. There certainly are indicators that the bias is artificial. For example, biochemistry majors are never observed to have a gender disparity. Similarly, Gen 341 and Orgo 351 and 352 show no gender bias for any of the majors. One similarity of these classes is that the exams are predominantly free-response rather than multiple-choice exams. Females typically do not perform as well on multiple-choice exams, which are based on a more algorithmic approach than free-response exams (Brownlow *et al.*, 2000). Similarly, active- and collaborative-learning experiences could be added, which may benefit female and racial minority students (Beichner *et al.*, 1999; Springer *et al.*, 1999). The historical trend in PHY 431 hints at the role that class size may play in gender disparity (Kokkelenberg *et al.*, 2008). As the size of the class increased, so did the difference in performance between the genders. Although class size is only a correlation in this example, it does remind us of the importance that this factor may play in the atmosphere of the classroom. It also suggests that the most efficient manner of "teaching" students may not be related to the most efficient student "learning" nor to retaining science majors.

LIMITATIONS OF THE STUDY

The models used in this study are not meant to predict individual students' performance; rather, they generalize the performances of a large group of students. The large university setting and large-enrollment class sizes at MSU provide further limitations (Kokkelenberg *et al.*, 2008). Smaller universities with small-enrollment classes could yield drastically different results due to a different student experience (Griffith, 2010). Another limitation specific to MSU is the use of multiple-choice exams in almost all large-enrollment courses, which makes the results less comparable to other universities that may use short-answer or essay exams.

CONCLUSION

Entering GPA was the best predictor found in the HLMs (β values of 0.315–0.662), followed by *Gender* (β values of -0.021 to -0.133) and *Major* (β values of 0.027 to -0.114), which

together explain 25–60% of the variance across the courses analyzed. Student involvement in *Honors College* or *LBC* and *Ethnicity* did not provide significant predictive ability to the models. A gender gap was consistently observed throughout a subset of students and could be the result of many factors. The gender variable may be acting as a proxy for the impact of negative signals that accumulate throughout a student's experiences, but proof of the source cannot be determined by this study. Further investigation of the specific sources of the performance differences needs to be done in order to fully understand and possibly correct the impacts of the leaky pipeline present in large-enrollment universities.

ACKNOWLEDGMENTS

The authors thank Dr. Mark Urban-Lurain for his input regarding the statistical methods used. L.R.C. thanks the Lyman Briggs College Undergraduate Research Fellows program for financial support. This work is based on work supported by the National Science Foundation under grants 0633222 and 1022754.

REFERENCES

- Astin AW (1993). *What Matters in College?* San Francisco: Jossey-Bass Publishers.
- Beichner R *et al.* (1999). Case study of the physics component of an integrated curriculum. *Am J Phys* 67, S16–S24.
- Beyer S (1999). Gender differences in the accuracy of grade expectations and evaluations. *Sex Roles* 41, 279–296.
- Boyer Commission (1998). *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*, Stony Brook: State University of New York at Stony Brook.
- Brownlow S, Jacobi T, Rogers M (2000). Science anxiety as a function of gender and experience. *Sex Roles* 42, 119–131.
- Cheng W, Ickes W (2009). Conscientiousness and self-motivation as mutually compensatory predictors of university-level GPA. *Person Individual Diff* 47, 817–822.
- Fox MF, Firebaugh G (1992). Confidence in science: the gender gap. *Soc Sci Q* 73, 101–113.
- Griffith AL (2010). Persistence of women and minorities in STEM field majors: is it the school that matters? *Econ Educ Rev* 29, 911–922.
- Inkelas KK, Soldner M, Longerbeam SD, Leonard JB (2008). Differences in student outcomes by types of living-learning programs: the development of an empirical typology. *Res High Educ* 49, 495–512.
- Kokkelenberg EC, Dillon M, Christy SM (2008). The effects of class size on student grades at a public university. *Econ Educ Rev* 27, 221–233.
- Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science* 330, 1234–1237.
- National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine (2007). *Rising above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*, Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology, Washington, DC: National Academies Press.
- National Research Council (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, Washington, DC: National Academies Press.
- Nowell A, Hedges LV (1998). Trends in gender differences in academic achievement from 1960 to 1994: an analysis of differences in mean, variance, and extreme scores. *Sex Roles* 39, 21–43.
- Ost B (2010). The role of peers and grades in determining major persistence in the sciences. *Econ Educ Rev* 29, 923–934.
- Rauschenberger MM, Sweeder RD (2010). Gender performance differences in biochemistry. *Biochem Mol Biol Educ* 38, 380–384.
- Seymour E, Hewitt NM (1997). *Talking About Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview Press.
- Solnick SJ (1995). Changes in women's majors from entrance to graduation at women's and coeducational colleges. *Indust Labor Relat Rev* 48, 505–514.
- Springer L, Stanne ME, Donovan SS (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. *Rev Educ Res* 69, 21–51.
- SPSS (2010). *SPSS Statistics for Windows (version 19.0)*, Chicago, IL: IBM.
- Stassen MLA (2003). Student outcomes: the impact of varying living-learning community models. *Res High Educ* 44, 581–613.
- Turner RC, Lindsay HA (2003). Gender differences in cognitive and noncognitive factors related to achievement in organic chemistry. *J Chem Educ* 80, 563–568.
- Wright R, Cotner S, Winkel A (2009). Minimal impact of organic chemistry prerequisite on student performance in introductory biochemistry. *CBE Life Sci Educ* 8, 44–54.