Performance, Prediction, and Preparedness: Do Biology-Major-Specific Courses Provide an Advantage?

Emily Weigel; Georgia Institute of Technology, Atlanta, GA
Juanita Pardo-Sanchez; University of Michigan, Ann Arbor, MI

Subject / Problem

To better instruct, biology education has focused much of its attention on shaping student experiences. Much of this effort has focused on increasing active learning and student-centered instruction, particularly in comparison to traditional lecture (Freeman et al., 2014). In such classrooms, students are tasked to relate and find meaning in the material taught (Hurney, 2012), and the evidence shows students report comparatively greater satisfaction and course averages in 'active' biology classes (Garcia et al, 2015; Hurney, 2012). A broader move toward successful active instruction across departmental course offerings has both decreased the practical differences between how majors and non-majors courses are taught, and sparked interest and lessened fears and anxieties non-STEM majors have towards STEM classes (Garcia et al, 2015). However, it is unclear whether differences exist *within* STEM majors; that is, whether non-Biology *STEM* majors (hereafter 'non-majors') have similar experiences to Biology majors when enrolled in identical coursework. Nonetheless, at many institutions, grouping by major is common for course offerings (often due to reasons not based in pedagogy, but rather staffing and classroom use), despite little evidence on how such grouping affects the student learning experience, particularly *within* STEM. **Here we explore the experiences of Biology majors and non-majors in identically taught courses and ask: *Is there an evidenced need for grouping by major based on exam performance, preparation, and predictions about one's exam success?***

If courses are taught identically, a main driver of different student experiences between majors and non-majors courses is related to the students themselves. Comparisons with other students, which are particularly amplified in classrooms that make learning visible, shape academic self-concepts as students compare their academic abilities to that of others in the class (Cooper et al., 2018). By reducing an introductory class to only majors, who then compare against other majors, the academic self-concept of a major student may be artificially determined by relative college preparedness rather than a students' true ability within the major. Negative feelings of one's performance could lead to a change in major early in a students' academic career when perceiving oneself to have less ability than the same performance would predict in a more general course. In fact, increased anxiety and concomitant decreases performance may impact persistence, as students disengage to attempt to avoid embarrassment in front of their peers (England et al., 2017). This is amplified in major-specific curricula if students are repeatedly grouped by major. Thus, efforts to retain students in majors by offering a major-specific section may actually backfire if not carefully crafted, particularly in early courses.

As part of the careful design of courses, departments must generally find not only additional space, but also the faculty to teach these specific courses. Similarly, sections that are grouped by major reduce the scheduling flexibility of students. This strain on departmental resources and students, without clear and compelling evidence to support the separate offerings, needs addressing. This paper therefore aims to investigate whether the use of majors- and non-majors sections of an identical introductory organismal biology course impacts student exam performance, score prediction, and preparedness dependent on course section taken.

Study Design

Students at a large, southeastern R1 university were recruited for study following IRB protocol #H17452. At this institute, an introductory-level organismal biology course is taught in two sections: a biology-majors-only section (total class size N=78, N=75 consenting for study, 60% female, 64% first-years), and a non-major-specific section of other STEM majors (e.g. Chemistry, Math, Computer Science, and Engineering majors; total class size N=103, N=98 consenting for study, 45% female, 38% first-years). Each section was team-taught in an active-learning format by the same two seasoned instructors for 50 minutes every Monday, Wednesday and Friday. The sections were taught identically one-hour apart using the same course materials. Students had an online pre-reading and associated set of 5 questions to prepare per class meeting, and weekly online home-works; feedback from these assessments, and in-class assignments, was given digitally within 2 hours of being due.

Throughout the semester, 4 in-person, 35-question multiple-choice exams with questions spanning all Bloom's levels were administered. After each exam, students were asked to fill out an 'Exam Wrapper' survey (Thompson, 2012) which included 5-item Likert-scale and short, open-ended questions regarding student's perceived anxiety level in preparing for the exam, how much time they spent studying (both in number of days and total number of hours), how satisfied they were with their studying, and their predicted score on the exam. Combined with the actual score on the exam, we were also able to analyze their predictive ability in assessing their exam scores by calculating the difference between the predicted score and their actual score on each exam. Results from the exam wrappers and test performance between the two classes were all analyzed using R (Team, 2017).

Analyses and Findings

To determine the suitability of standard parametric tests for our analyses, a Shapiro Wilks test was conducted on each variable to determine its normality in addition to using a visual assessment. Each variable measured failed the Shapiro Wilks test (all $p<0.05$), indicating non-normality and inappropriateness of standard tests. In response to this, we first transformed the data: If the data was skewed left, then the square root of the response was taken. If the data was skewed right, then the log was taken of the data. Specifics of the transformation prior to model fitting are indicated below. To confirm variable transformation did not significantly impact the outcomes of our models, we also ran the same models without transformation; we achieved similar results, so the most appropriate, transformed models are reported below.

Base linear mixed effects models were created by using exam number and class as fixed effects. Gender, student, and year in college were used as random effects to account for important differences between students and classroom populations. The sole interaction initially included was that between exam number and class. The best fit model was deduced using AIC in all cases by stepwise removal predictors and their interactions. The following linear mixed effects models were used to look at the predictive effects of these variables on:

1. Exam scores
2. Anxiety levels reported by students
3. Number of days spent preparing for the exam
4. Satisfaction levels with their studying
5. Number of hours spent preparing for the exam
6. Exam score prediction
7. The difference between their predicted and actual exam score

The Shapiro Wilks test for the exam scores determined that the scores were not normally distributed ($p<0.05$). Using a visual assessment, it was determined that the scores were skewed to the right. To correct for this, the log of the exam scores was taken. The best fit model for the effect of class and exam on exam scores only used class as a fixed effect. A summary of the model resulted with no significant difference in exam scores between classes ($-4.166 \leq t \leq 3.098$, df = 172, N = 173). A summary of the exam scores can be found in Table 1.

Table 1. Class average and SEM for each exam. A Kruskal-Wallis test was conducted between exam scores separated by class and exam resulted in no significant difference between either classes in any of the exams (p > 0.05, df=172, N=173).

Class	*Exam 1*	*Exam 2*	*Exam 3*	*Exam 4*
Non-Majors	80.92±1.47	77.85±1.85	84.51±1.66	77.72±1.87
Majors	78.78±1.41	78.81±.1.97.	84.48±1.95	73.97±1.58

Students reported their anxiety levels by selecting a level of anxiety most closely related to theirs (very anxious to very calm, lowest to highest on a 5-point Likert scale). This categorical data was transformed into factors for statistical analysis. The best fit model for anxiety included class and exam number as fixed effects. No significant difference between anxiety levels was found ($-2.024 \leq t \leq 0.149$, df = 172, N = 173).

Students were also asked to report the number of days they spent preparing for each exam. This categorical data was also transformed into factors. The best fit model analyzing the number of days students spent studying included exam number and class as fixed effects. Classes did not differ in their days spent preparing for exams ($-3.186 \leq t \leq 1.186$, df = 172, N = 173).

Student satisfaction levels with their studying were recorded in five different categorical levels (very dissatisfied to very satisfied, lowest to highest on a 5-point Likert scale) which were transformed to factors. The lowest option corresponded to the number one and the highest option corresponded to the number five. The best fit model for satisfaction levels as reported by students included class and exam number alone. No significant difference between studying satisfaction levels was found ($-1.034 \leq t \leq 2.902$, df = 172, N = 173).

Students were asked to quantify the number of hours spent preparing for each exam. This data was skewed to the left determined visually and with a Shapiro Wilks test ($p<0.05$). To correct for this, the square root was taken of the predicted exam scores. The best fit model included exam and class as fixed effects. No significant difference was found between number of hours studying for the exam and the class or exam ($1.849 \leq t \leq 28.496$, df = 172, N = 173).

After exams, students were asked to estimate the score they received. This data was skewed to the left determined visually and with a Shapiro Wilks test ($p<0.05$). To correct for this, the square root was taken of the predicted exam scores. The best fit model found for this response included exam number and class as fixed effects. Predictions were not significantly different between classes for any of the exams ($-1.928 \leq t \leq 2.227$, df = 172, N = 173).

The difference and absolute value between the observed exam score and the students' expected score was taken; this was to assess the potential accuracy of exam score predictive ability and its potential change both across the semester and between courses. The best fit model for this case included only exam number, and which was not found to be significant ($-0.513 \leq t \leq 1.968$, df = 172, N = 173).

Contribution

While based on a single instance of two identically-taught courses, this study provides some evidence that, within STEM majors, biology-majors may not see increased benefit to sections offered solely to them in terms of their performance, preparation (and feelings around such preparation) and their metacognitive predictions about how such preparations may have impacted performance. Future studies should explicitly investigate how other factors not measured here, such as potential differences in credit hours taken or sense of belonging in the courses, may affect other aspects of student experience, particularly in ways that impact persistence in the major.

In accordance with previous studies, active learning lead to academic achievement of students in introductory biology courses (Hurney, 2012), as evidenced by relatively high averages inclusive of standard deviation on exams that challenged students across Bloom's levels However, despite high performance on exams, major and non-major students reported similar levels of anxiety caused by exams, and this anxiety did not shift decrease across the semester, even as students grew familiar with the exam format. This result indicates that regardless of the students' relation to the course topic, exam anxiety persists in similar forms and may be hard to alleviate, although being a course related to your major does not appear to alleviate nor increase one's anxiety relate to the course. Given that this course used techniques, such as clickers, to allow for anonymous answering of questions to reduce anxiety (England et al., 2017), it is perhaps pertinent to consider teaching students techniques specific to reducing anxiety while specifically preparing for exams, rather than in generally engaging face-to-face in a course.

Exam wrappers are a tool for teaching reflective metacognitive skills, particularly cognizance related to effective study techniques, and they are touted to help students set better goals for the next exam (Gezer-Templeton et al., 2017). However, this study found no significant difference between majors and non-majors across four exams in the number of hours and days they studied for the exam nor in their satisfaction with their studying. Indeed, studies show that students vary in metacognitive abilities, particularly with age and practice (Li et al., 2016), and we saw more variation within than between courses. Furthermore, as students showed relatively constant study behaviors over time, it is possible that past study techniques, even when ineffective, are 'ingrained' from K-12 experiences; for this reason, future studies should consider the instruction and feedback students received in how to study. As we saw students across all levels of achievement reluctant to switch techniques, some instruction on alternative study strategies or how to form study groups, could be incorporated. This is particularly important as major-specific, cohort models are touted as ways to build retention and skills via connections with other students (although evidence is lacking). Simply, we may need to 'correct for' the comfort of certain techniques and social pressure to adopt them, even when subpar; the students simply may not know how to go about studying another way unless we show them.

Students often overestimate their predictions of exam scores, even after multiple exams (Foster et al., 2017). Consistent with prior studies, we found that exam score predictions did not improve over time and were equally over-inflated relative to actual exam score for majors and non-majors. Although we found across courses higher-achieving students tended to be more accurate (as in Bol et al., 2005), we did not find that these students grew any more accurate across the semester. Knowing that poor estimation of exam score does not depend on major, but rather more generally on level of achievement, instructors may use predictive abilities to target students who may need additional assistance; this may be particularly pertinent to help students who assume incorrectly that they will 'automatically' perform better in major courses simply because they have declared a particular major.

General Interest

From this study, we have found that biology and non-biology stem majors experience active learning introductory biology classrooms in similar ways with respect to exam performance, preparation, and predicted performance. Our evidence suggests that courses restrictive by major (or perhaps in a K-12 system, by track) may not be as beneficial as grouping students by direct assessment of biology or metacognitive skills. As there appears to be more uniformity in response within STEM majors, this also suggests that separating introductory biology classes by specific major holds no direct achievement benefit if both are taught utilizing active learning. The release from requiring separate sections may relieve departmental burden from finding additional teaching staff and also help students to stay on track by allowing for more scheduling flexibility in the courses a student may take in a given semester. Our data also suggest that elements of test anxiety and study strategy may be difficult to change within a semester; this calls for a deliberate, curriculum-wide, longer-term focus on helping students to develop self-assessment and effective study skills, particularly for lower-performing students.

References

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73*(4), 269-290. doi:Doi 10.3200/Jexe.73.4.269-290

Cooper, K. M., Krieg, A., & Brownell, S. E. (2018). Who perceives they are smarter? Exploring the influence of student characteristics on student academic self-concept in physiology. *Advances in Physiology Education, 42*(2), 200-208. doi:10.1152/advan.00085.2017

England, B. J., Brigati, J. R., & Schussler, E. E. (2017). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PLoS One, 12*(8). doi:ARTN e0182506Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: the role of memory for past exam performance in student predictions. *Metacognition and Learning, 12*(1), 1-19. doi:10.1007/s11409-016-9158-6

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111*(23), 8410-8415. doi:10.1073/pnas.1319030111

Garcia, R., Rahman, A., & Klein, J. G. (2015). Engaging Non—Science Majors in Biology, One Disease at a Time. *The American Biology Teacher, 77*(3), 178-183. doi:10.1525/abt.2015.77.3.5

Gezer-Templeton, P. G., Mayhew, E. J., Korte, D. S., & Schmidt, S. J. (2017). Use of Exam Wrappers to Enhance Students' Metacognitive Skills in a Large Introductory Food Science and Human Nutrition Course. *Journal of Food Science Education, 16*(1), 28-36. doi:10.1111/1541-4329.12103

Hurney, C. A. (2012). Learner-centered teaching in nonmajors introductory biology: the impact of giving students choices. *J Microbiol Biol Educ, 13*(2), 133-141. doi:10.1128/jmbe.v13i2.458

Li, W., Ji, H., Li, F., Li, P., Zhang, Y., & Li, X. (2016). Metacognitive control: Shifting from habitual to agenda processes on item selection during study in elementary school children. *Br J Educ Psychol, 86*(4), 657-670. doi:10.1111/bjep.12130

Team, R. C. (2017). R: A Language and Environment for Statistical Computing (Version 3.4.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Thompson, D. R. (2012). Promoting Metacognitive Skills in Intermediate Spanish: Report of a Classroom Research Project. *Foreign Language Annals, 45*(3), 447-462. doi:10.1111/j.1944-9720.2012.01199.x

Yip, M. C. W. (2007). Differences in Learning and Study Strategies between High and Low Achieving University Students: A Hong Kong study. *Educational Psychology, 27*(5), 597-606. doi:10.1080/01443410701309126