

Background

Prior literature has shown disparities in introductory biology exam scores between students with different social identities such as gender and socio-economic status (Eddy & Brownell 2016, Wright et al. 2016). Exams are often high-pressure situations that could trigger stereotype threat for students holding social identities associated with negative stereotypes (Spencer et al. 1999). Offering optional exam retakes could be one way to reduce the pressure on student performance and reduce observed grade disparities. Moreover, repeated testing has been shown to enhance student learning (Roediger & Karpicke, 2006). Therefore, offering students an opportunity to retake exams might improve student learning and reduce grade disparities. We examined two introductory biology courses that offered optional exam retakes and asked:

1. Who takes exam retakes?
2. What is the effect of optional exam retakes on student grades?
3. What is the effect of optional exam retakes on student grade gaps by social identities such as gender, race/ethnicity, and number of hours worked?

Methods

Data collection: This study was conducted at a large nonselective public university in the Southwestern US. We collected data on student performance on exams from two large introductory biology courses that constitute the introductory biology sequence for biology majors in Fall 2019 (n= 764) and Spring 2020 (n= 614) that offered optional exam retakes to all students. We also collected demographic information through the registrar office and a student survey.

Both courses had three high-stakes multiple-choice exams each. In Fall 2019, the retakes were offered outside of regular class period with multiple time slot options in a day. However, in Spring 2020, the retakes were offered during class periods before the transition to remote learning due to COVID-19. In both semesters, the higher of the two exam scores (i.e. first attempt and retake) were counted towards the final grade.

Data analysis: We used logistic regressions to assess which variables influence the likelihood of a student retaking an exam. We used total student exam score on first attempts, gender, race/ethnicity, Pell-grant eligibility, and number of hours that students worked a job during the semester as predictors. Next, we used paired t-tests to assess if retaking exams improved student exam scores. Finally, to assess whether retaking exams reduced grade disparities, we ran a multiple regression with the total exam score students received in the course as the outcome while controlling for total exam score on first attempts.

Table 1. Student demographics. Note that we do not have information on number of hours that students worked a job per week for 25 students in Course 2.

	Course 1	Course 2
Gender	N (%)	N (%)
Women	520 (68.1)	421 (68.6)
Men	244 (31.9)	193 (31.4)
Race/ethnicity		
Asian	108 (14.1)	100 (16.3)
Black/African American	39 (5.1)	34 (5.5)
Hispanic/Latinx	193 (25.3)	139 (22.6)
International	26 (3.4)	24 (3.9)
Multiracial	45 (5.9)	36 (5.9)
White	353 (46.2)	281 (45.8)
Work		
Did not work	439 (57.5)	405 (68.8)
1-20 hours	236 (30.9)	105 (17.8)
More than 20 hours	89 (11.6)	79 (13.4)
Pell grant eligibility		
No	512 (67.0)	421 (68.6)
Yes	252 (33.0)	193 (31.4)

Results

1. Students with lower scores on first attempts retake exams more often

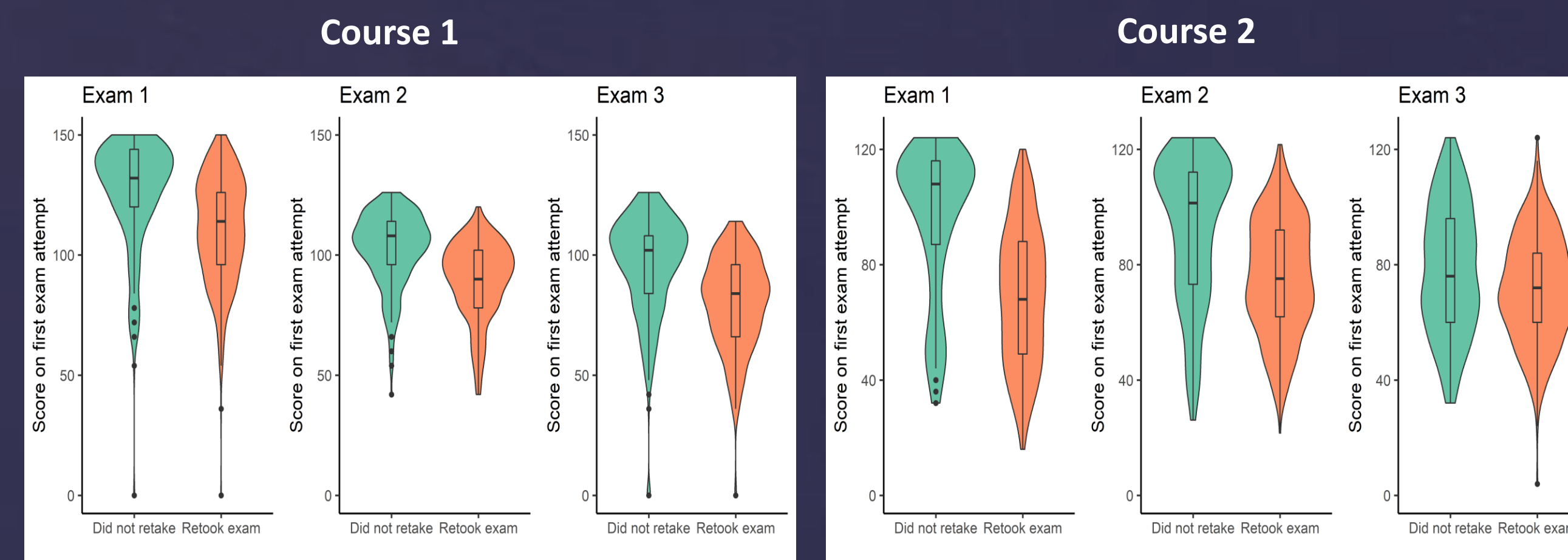


Figure 1. Score on first attempts for students that did not retake an exam (in green) and those that retook an exam (in orange). Violins show the distribution of the data, grey bars in the middle of each violin indicates the interquartile range and black lines in the middle of these bars indicate the median, black lines stretched out from the bar indicate first/third quartile \pm 1.5 interquartile range.

2. There are demographic differences in participation in optional exam retakes

Table 2. Odds ratio estimates with 95% confidence intervals (indicated by 2.5% and 97.5%) of students retaking exams from logistic regressions. Odds ratios lower than 1 indicate lower likelihood of retaking exams compared to reference groups and vice versa. Bolded blue font indicates significant differences (i.e. $p < 0.05$). Students that scored higher on first attempts were less likely to retake exams. Moreover, students that work, Black/African American students, and Hispanic/Latinx students were less likely to retake exams.

	Course 1			Course 2		
	Odds Ratio	2.5%	97.5%	Odds Ratio	2.5%	97.5%
Exam total on first attempts	0.60	0.55	0.67	0.61	0.54	0.69
Gender (ref. Women): Men	1.03	0.85	1.24	0.86	0.68	1.08
Pell grant eligibility (ref. No): Yes	1.40	1.14	1.71	1.05	0.81	1.37
Race/ethnicity (ref. white):						
Asian	1.83	1.40	2.39	1.00	0.73	1.37
Black/African American	0.76	0.50	1.15	0.56	0.35	0.90
Hispanic/Latinx	0.76	0.61	0.95	0.90	0.67	1.22
International	2.26	1.36	3.85	1.15	0.66	2.09
Multiracial	0.79	0.54	1.14	1.02	0.63	1.70
Work (ref. Did not work):						
1-20 hours	0.81	0.66	0.98	0.94	0.70	1.27
More than 20 hours	0.76	0.58	1.00	0.72	0.52	1.01

3. Students score higher on exam retakes compared to first attempts

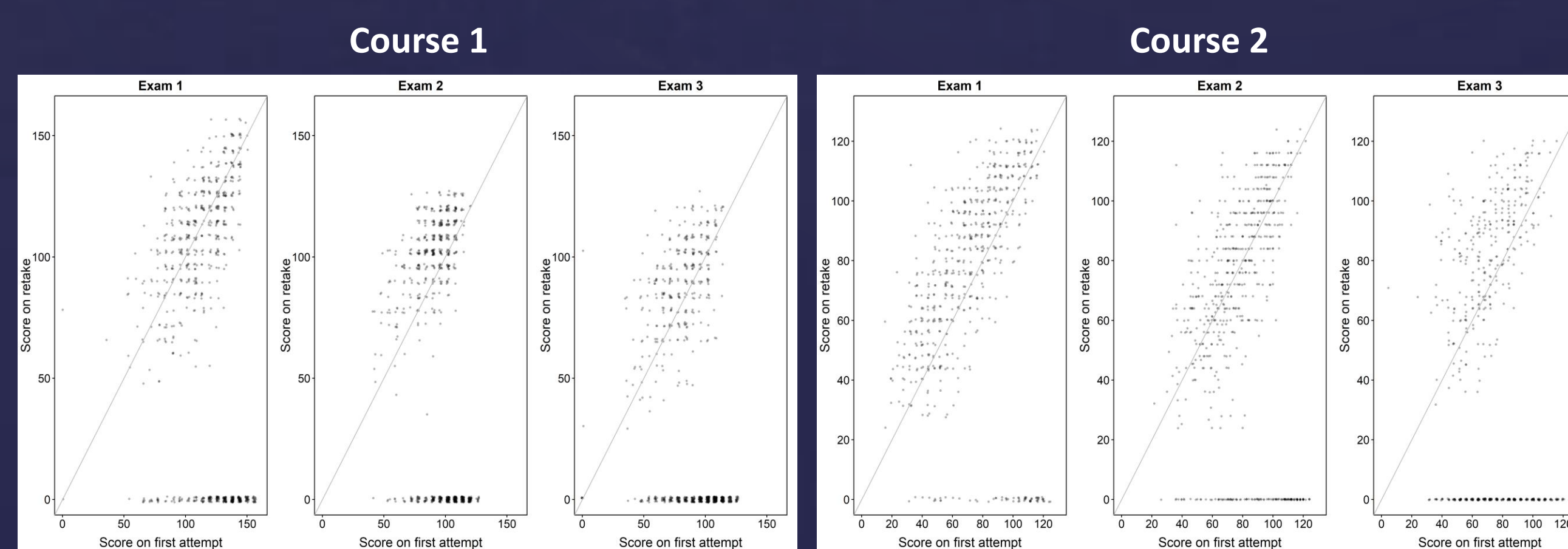


Figure 2. Score on first attempts and optional retakes for students in introductory biology courses. Each point indicates a student, points on the diagonal line show students that scored the same on first attempts and retakes. Points above the line indicate students that did better on retakes. (paired t-tests, all p-values except Course 1 Exam 1 were less than 0.05).

4. Exam retakes have no effect on disparities in scores between students with different social identities

Table 3. Regression slope estimates for actual score received after taking exam retakes into account while controlling for total exam score on first attempts. Bolded blue font indicates p-values less than 0.05 for the slope estimate. Exam retakes did not reduce demographic disparities in exam scores, in either of the two courses. However, the score disparities between white, and Black/African American and International students increased after retakes in Course 1 and score disparities between students that did not work and those that worked more than 20 hours increased in Course 2.

	Course 1 (Slope \pm SE)	Course 2 (Slope \pm SE)
(Intercept)	0.01 \pm 0.03	0.03 \pm 0.03
Exam score on first attempts	0.92 \pm 0.01	0.92 \pm 0.02
Gender (ref. Women): Men	0.02 \pm 0.03	-0.04 \pm 0.03
Pell grant eligibility (ref. No): Yes	0.03 \pm 0.03	-0.03 \pm 0.03
Race/ethnicity (ref. white):		
Asian	0.05 \pm 0.04	0.06 \pm 0.04
Black/African American	-0.13 \pm 0.06	-0.06 \pm 0.07
Hispanic/Latinx	-0.04 \pm 0.03	-0.04 \pm 0.04
International	-0.15 \pm 0.08	0.03 \pm 0.08
Multiracial	-0.01 \pm 0.06	0.03 \pm 0.07
Work (ref. Did not work)		
1-20 hours	-0.02 \pm 0.03	-0.01 \pm 0.04
More than 20 hours	-0.06 \pm 0.04	-0.10 \pm 0.04

Conclusions

- Students with lower scores tended to retake exams more often, but the likelihood of retaking exams varied across demographic groups, even after taking score on original exams into account.
- Students who work were less likely to retake exams than their counterparts. One potential explanation is that preparing for the second exam and retaking exams are time consuming, making it difficult for these students to take advantage of optional exam retakes.
- There were inconsistent differences in likelihood of retaking exams by race/ethnicity. Asian and International students were more likely to retake exams in Course 1, but not in Course 2. Hispanic/Latinx students were less likely to retake exams in Course 1 while Black/African American students were less likely to retake exams in Course 2. Differences in when the retakes were scheduled might contribute to these differences.
- Broadly, students do better on retakes than the first attempts suggesting that optional exam retakes might be a useful tool in class to improve student performance and learning.
- However, optional exam retakes don't reduce observed demographic disparities in total score across first attempts and might even slightly increase them. Differences in the likelihood of retaking exams partially explains this and raises the question of whether exam retakes should be optional. Another potential explanation is that high-stakes exams remain high pressure situations despite optional exam retakes and therefore, stereotype threat continues to operate.
- One major limitation of the current study is that Course 2 data was collected in Spring 2020, when the semester and lives were heavily disrupted due to COVID-19.

Literature cited

1. Eddy, S. L. & Brownell, S. E. (2016). *Physical Review Physics Education Research*, 12(2), 020106.
2. Wright, C. D., Eddy, S. L., Wenderoth, M. P., Abshire, E., Blankenbiller, M., & Brownell, S. E. (2016). *CBE—Life Sciences Education*, 15(2), ar23.
3. Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, 35(1), 4-28.
4. Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, 1(3), 181-210.

Contact info

Email: ksupriya@asu.edu

Website: <http://sebbbers.wixsite.com/biology-ed-lab>; <http://ksupriya.weebly.com/>

ASU_BioEdLab supriya_sci

Funding source: NSF GR30875