Examining the Impact of Case Studies on Student Learning, Interest, Motivation, and Belonging in Undergraduate Human Physiology

M.M. Pekary, J.K. Jellyman, M.T. Giang, and P.M. Beardsley

Corresponding Author: mmhays@cpp.edu

HAPS Educator. Vol 25 (2), pp. 30-43. Published August 2021.

https://doi.org/10.21692/haps.2021.023

Pekary MM, Jellyman JK, Giang MT, and Beardsley PM (2021). Examining the Impact of Case Studies on Student Learning, Interest, Motivation, and Belonging in Undergraduate Human Physiology. *HAPS Educator* Vol 25 (2), pp. 30-43. https://doi.org/10.21692/haps.2021.023

Examining the Impact of Case Studies on Student Learning, Interest, Motivation, and Belonging in Undergraduate Human Physiology

M.M. Pekary, PhD[1]*, J.K. Jellyman, PhD[1], M.T. Giang, PhD[2], P.M. Beardsley, PhD[1,3]

[1] Biological Sciences Department, California State Polytechnic University at Pomona, Pomona, CA, mmhays@cpp.edu, jkjellyman@cpp.edu, pmbeardsley@cpp.edu

[2] Psychology Department, California State Polytechnic University at Pomona, Pomona, CA, mtgiang@cpp.edu

[3] The Center for Excellence in Mathematics and Science Teaching, California State Polytechnic University at Pomona, Pomona, CA, pmbeardsley@cpp.edu

* **Corresponding Author**: Mackenzie Pekary, Department of Biological Sciences, California State Polytechnic University at Pomona, 3801 W Temple Ave, Pomona, CA 91768

Abstract

Human physiology is a foundational course for future health practitioners. Previous research suggests case studies may improve students' critical thinking and comprehension, but little is known about how case studies affect student interest, motivation, and belonging. The current quasi-experiment investigates the impact of integrating case studies into an otherwise lecture-based human physiology course, compared to traditional lecture-based teaching. We hypothesized that case studies would improve all student outcomes, especially for students from underrepresented groups. Results showed that students in the case study section received higher overall grades and scores for related questions on three exams, combined higher level Bloom's questions on those exams, interest, motivation, and belonging to the course when compared to at least one, and in most cases both, of the comparison sections. We found few significant interactions between course section and student demographic groups, though improvements in final grades appear to be driven by improved grades for women. https://doi.org/10.21692/haps.2021.023

Key words: case studies, human physiology, active learning

Introduction

Lower division human physiology courses are an important gateway to careers in the health science field as they help prepare students for upper division classes and taking health profession entry exams (Cliff and Wright 1996; Smee and Cooke 2018). In these courses, students learn a broad range of content, including interactions spanning almost every system in the body. However, students must also be able to apply their knowledge and draw connections amongst concepts (Cliff and Wright 1996; Ediger 2017; Smee and Cooke 2018). The ability to apply knowledge and draw connections amongst concepts requires higher levels of thinking. The large volume of content covered in the course can conflict with other important learning objectives, especially those involving higher-level thinking skills that ask students to apply their knowledge to medically-related examples. Smee and Cooke (2018) believed students were too focused on remembering facts rather than understanding them and applying them.

Improving teaching in human physiology courses can help students from a broad range of backgrounds and racial and ethnic identities be successful. Improving learning and success for all students is critical for diversifying the future healthcare work force. More diverse teams are widely recognized as being more successful on a range of metrics (Cheruvelil et al. 2014). Better enabling the success of all students, especially those from minoritized groups, may have direct benefits for those receiving care as patients may be more responsive to healthcare workers with a similar background or identity. For example, a study conducted by Alsan et al. (2018) found that Black men were more likely to undergo preventative health care screenings and services if their doctor was also Black. However, according to the Association of American Medical Colleges, in 2018 only 5% of active physicians identified as Black or African American (AAMC 2018). Unfortunately, lower-level university science, technology, engineering and mathematics (STEM) courses generally have achievement or opportunity gaps in which students from underrepresented and lower socio-economic status backgrounds receive lower grades and have a higher probability of failing (Theobald et al. 2020). More specifically, according to student success statistics provided by the university at which our study was conducted, there is a 14.5% achievement gap in final grades between underrepresented minority (URM) students and non-underrepresented minority (non URM) students taking the introductory human physiology course. Additionally, there is a 7.5% achievement gap in final grades between first generation (first gen) students and non-first generation (non-first gen) students taking the course.

continued on next page

Integrating case studies into the teaching of human physiology may be a promising start to improving student learning (Birk et al. 2019; Cliff and Wright 1996; Smee and Cooke 2018). Case studies are student-centered activities that contextualize concepts by giving students realistic practical problems to solve while using higher level thinking (Allchin 2013; Herreid 1994; Herreid 1997). Case studies have three key components; an engaging narrative containing contextualized information, questions targeting learning objectives, and class discussion of students' solutions. (Cliff and Wright 1996; Herreid 1994; Tomey 2003; Wilcox 1999). Cliff and Wright (1996) found that the use of case studies increased student test scores in human physiology lecture classes. Researchers believe this improvement in learning is due to the fact that case studies contextualize information making it easier to visualize and more relatable to the learner and by reinforcing key concepts by focusing on key objectives, (Bonney 2015; Cliff & Wright 1996; Freeman et al. 2014; Herreid 1994; Smee and Cooke 2018). According to Allchin (2013), "contextualizing the learning contributes both to student motivation and to the making of meaning (construed by many educators as central to functional memory and effective learning)."

Case studies may also be useful for decreasing achievement gaps. Case studies almost exactly match the definition of active learning used by Freeman and colleagues (2014) because they emphasize higher order thinking and students learn by participating in activities and discussions in class rather than just listening to lectures. Research has shown that active learning disproportionately benefits underrepresented groups in STEM, possibly because it provides more opportunities to engage with material (Birk et al. 2019; Theobald et al. 2020).

Case studies may simultaneously increase student achievement (Cliff and Wright 1996) and, because it involves active learning, decrease achievement gaps by improving students' interest and sense of belonging, which are key aspects of student motivation to learn (Theobald et al. 2020). Motivation has been shown to be a key indicator of student performance in a course (Getty et al. 2017). A widely used model of motivation is the expectancy value theory (Eccles 1983), which has recently been expanded to include cost (Getty et al. 2017). Expectancy refers to how well a student expects to do in the course and value refers to how important the course is to a student (Getty et al. 2017; Barron et al. 2017; Eccles 1983). According to Eccles (1983), expectancy and value can have a large effect on student performance. Cost refers to what barriers might prevent a student from succeeding in the course (Getty et al. 2017; Barron et al. 2017). It is a newer factor to the model, but researchers believe that it is important to take into consideration because it measures factors that may decrease a student's motivation to learn, and evidence suggests it is a separate construct from value (Getty et al. 2017). Additional psychosocial factors are also important. According to the National Academies of Science, interest and sense of belonging are two of the eight intra- and interpersonal competencies related to achievement in undergraduate education (National Academies of Sciences 2017). Research compiled by the National Academies of Science collectively supplied evidence that a student's sense of belonging in college is mutable and that a higher sense of belonging is correlated with greater achievement academically (National Academies of Sciences 2017). Additionally, rigorous studies using interventions to increase the relevance students see in the course content differentially benefitted students from underrepresented groups (Canning et al. 2018). Case studies may improve student interest and possibly belonging because the stories spark triggered situational interest where the narrative in the story initially draws the student in, while the group work to complete the problems fosters an ideal situation for maintained situational interest to take hold (Hidi and Renninger 2006). Additionally, contextualization makes the course content relatable and relevant to students. "Real-life" connections may especially help cultivate interest in females and URM students (Allchin 2013).

While case studies hold promise, much remains to be learned about the extent to which this pedagogical approach achieves its aims and the mechanism by which the method may help students achieve these aims. More empirical research is needed to begin to address these claims. Therefore, our research questions are the following:

1. Among students enrolled in a sophomore-level human physiology course, to what extent do learning gains, interest, motivation, and belonging differ between students who have received case study-based instruction and those who have received traditional lecture-based instruction?
2. To what extent do learning gains, interest, motivation, and belonging outcomes vary for students from different demographic groups (first-generation students versus non first-generation students, underrepresented students versus not, and females versus males)?

Methods

Experimental design

We designed a quasi-experiment using three sections of an undergraduate course in human physiology. One section was taught using case studies (treatment section; Table 1) and two additional sections of the same course, each taught by different instructors, were taught using "business as usual" approaches (comparison sections; Table 1). The curriculum was taught in three five-week blocks, each of which was assessed using an exam (Figure 1). Instructors of the comparison sections primarily used a passive lecture format with heavy reliance on presentation slides. Course content was aligned with the textbook content and the instructors did not use case studies. The instructor of the treatment section mainly used passive lectures that relied on presentation slides but also incorporated three case studies. The complexity and format of the science content was similar in all lecture sections. The

continued on next page

instructors of each section voluntarily agreed to take part in the study. The study was approved by the Institutional Review Board (Protocol #19-236). Students consented to take part in the study without any knowledge of the study's objectives or hypotheses. To incentivize participation, students received 2% of extra credit at the end of the semester in return for completing two surveys (Figure 1).

Treatment Group	Previously taught this course	Used a learning assistant	Gave study guides
Treatment	No	Yes	No
Comparison 1	Yes	Yes	Yes
Comparison 2	Yes	No	Yes

Table 1. *Instructor differences for each human physiology course section*

Differences between course sections

Due to the quasi-experimental nature of the study, differences between course sections other than those being studied were inevitable (Table 1). Every effort was made to control and address various differences between the course sections. The treatment section was taught by an instructor new to teaching the course in order to mitigate instructor experience from playing a larger role. The treatment and comparison 1 sections were scheduled to be held twice a week for an hour and 50 minutes each while the comparison 2 section was scheduled to meet three times a week for 50 minutes each. However, with the transition to online learning necessitated by the COVID-19 pandemic, all instructors switched to releasing lecture material once a week.

Three exams given throughout the semester made up the bulk of students' final grades in all course sections. Each instructor wrote their own exam, so the total number of questions varied. All three exams were multiple choice with the exception that for exam 1 which was given in a face-to-face setting, one of the comparison instructors included a few free response questions. The only uniform questions on each exam were the researcher developed questions, which were the ones we used to assess student learning (see the data collection instruments section). Other factors included in final grades were short quizzes and homework assignments worth small amounts of points in the comparison sections and case study worksheets in the treatment section. Students in the treatment section were given credit for completing the case studies and participating in the discussion when going over the solutions. Students were not graded on the correctness of their answers.

The treatment section and comparison 1 section both had a learning assistant assigned to their course (Table 1). Learning assistants are undergraduates who serve as supports for the instructors and students during class and who also offer individualized tutoring.

Both comparison instructors gave students a study guide for all exams (Table 1). The study guides contained long lists of topics and learning objectives for students to review as they prepared for exams.

Student population

Student subjects in this study were voluntarily recruited from three sections of a 16-week introductory undergraduate sophomore-level human physiology course at a large, urban, public, Masters-granting Hispanic Serving Institution in California during the spring semester of 2020. A total of 242 students were enrolled in the course during this time; 224 consented to participate in the study (Table 2). Greater than 85% participated and completed all the assessments (Table 2).

	Treatment	Comparison 1	Comparison 2
Completed Pre-survey	77 (95.1%)	52 (86.7%)	94 (93.1%)
Completed Exam 1	77 (95.1%)	52 (86.7%)	95 (94.1%)
Completed Exam 2	77 (95.1%)	51 (85.0%)	95 (94.1%)
Completed Exam 3	76 (93.8%)	51 (85.0%)	94 (93.1%)
Completed Post-survey	71 (87.7%)	51 (85.0%)	87 (86.1%)
Note. One student consented to participate but did not fully complete the pre-survey.			
Treatment n=81, comparison 1 n=60, comparison 2 n=101.			

Table 2. *Percent participation in study at different stages in research*

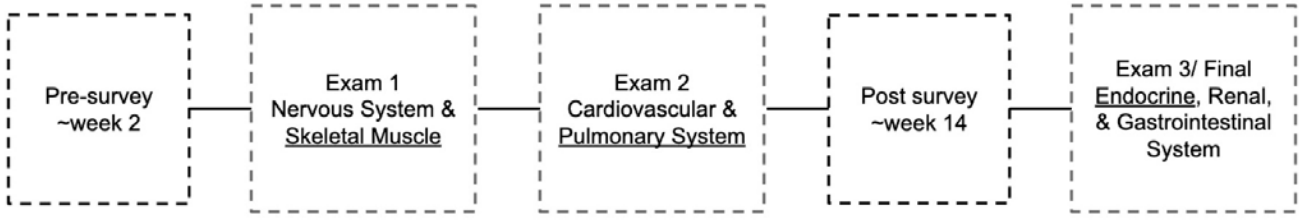

Figure 1. *Timeline of the study. The underlined topics on the exams were covered in the treatment section using case studies.*

continued on next page

Student demographic data and grades at the end of the semester were obtained from the Registrar's Office of the university. Of the students included in the study, 149 (66.5%) were female, 74 (33%) were male, and 1 (0.4%) did not supply identification. Eighty students self-identified as an under-represented minority by race or ethnicity (URM) (35.7%), 112 (50%) identified as non-URM and 32 (14.3%) chose not to identify. In addition, 104 students (46.4%) were the first in their family to either attend college or seek a degree (first generation, or first gen), 85 students (37.9%) were non-first generation, and 35 (15.6%) chose to not respond. See table 3 for details on demographic group breakdown by course section.

Case study learning activities

Relevant, freely available case studies from the National Center for Case Study Teaching in Science (NCCSTS; http://sciencecases.lib.buffalo.edu/cs/) were used as active learning activities to support the teaching of skeletal muscle physiology (All or Nothing: A Case study in Muscle Contraction, Neumann et al. 2016), pulmonary physiology (Asthma Attack, Leavitt 2018) and endocrine physiology (Muscleman: A Surprising Case of Shrinkage, Schillo 2012) in the treatment section. The case studies were given after students had reviewed the topic in a lecture format so students could practice applying what they learned. Time devoted to lecture in the case study section was decreased to give students time to complete and review the case study.

The case study on muscle physiology was conducted for ~30 minutes during a face-to-face class meeting. The case study questions were completed in small discussion groups of ~three students. The groups were formed by students partnering up with others sitting near them. The instructor was available to answer questions, but ultimately students were responsible for working through the questions in the case study within their groups. Solutions to the case study questions were delivered using an instructor-facilitated discussion at the end of the class period (Herreid 2005; Murray-Nseula 2011).

On March 19, 2020, the Governor of California issued an Executive Order and Public Health Order that directed all Californians to stay home except to go to an essential job or to shop for essential needs to mitigate the threat of the COVID-19 pandemic. As a result, all university courses switched to an online teaching mode. Due to this unexpected transition to online learning, the case studies on pulmonary and endocrine physiology were completed as asynchronous assignments in the treatment section. The discussion board forum on the course learning management system was used to facilitate student discussion. Students were required to either post a question or answer another student's question on the discussion board to receive full credit for completing the case study assignment. The case study solutions were delivered during a synchronous, instructor-led discussion that occurred via videoconference. Neither the instructor's written solutions, nor the published case study solutions were posted online at any time in accordance with the requirements of the NCCSTS.

Data collection instruments

The pre-survey contained instruments that measured students' interest (Linnenbrink-Garcia et al. 2010), motivation (Barron et al. 2017), and sense of belonging to the field of health science (Walton et al. 2015). We assessed student content knowledge of muscle, pulmonary, and endocrine physiology using a researcher-developed pre-survey prior to starting the course. The pre-survey contained four questions on skeletal muscle physiology, five questions on pulmonary physiology, and six questions on endocrine physiology. These questions were designed to correlate with the learning objectives for each topic. The post-survey was administered towards the end of the semester, just prior to students taking their final exam. The post-survey contained the same instruments as

Demographic Group	Treatment	Comparison 1	Comparison 2	Total	% of dataset
Female	52 (67.5%)	34 (65.4%)	63 (66.3%)	149	66.5%
Male	25 (32.5%)	18 (34.6%)	31 (32.6%)	74	33.0%
Sex Not Provided	0 (0.0%)	0 (0.0%)	1 (1.1%)	1	0.5%
URM	23 (29.9%	22 (42.3%)	35 (36.8%)	80	35.7%
Non-URM	42 (54.5%	24 (46.2%)	46 (48.4%)	112	50.0%
URM Status Not Provided	12 15.6%	6 (11.5%)	14 (14.7%)	32	14.3%
First Gen	32 (41.6%)	28 (53.8%)	44 (46.3%)	104	46.4%
Non-First Gen	32 (41.6%)	17 (32.7%)	36 (37.9%)	85	38.0%
First Gen Status Not Provided	13 (16.9%)	7 (13.5%)	15 (15.8%)	35	15.6%

Table 3. *Participation in study by demographic group*

continued on next page

the pre-survey minus the content knowledge questions and with additional instruments that measured students' sense of belonging to the course (Walton et al. 2015) and students' self-assessment of their learning gains (SALG). The student perception of learning gains instrument was developed using a modified form of the SALG instrument used by Bonney (2015) and researcher-developed questions. The modified SALG instrument used a 5-point Likert scale and included questions 11-16 from Bonney (2015). In addition, an "attending or listening to instructor's lectures" option was added to each question category. The researcher developed questions asked students to rate, using a 5-point Likert scale, the extent to which they believed they understood each of the three subject areas before and after taking the course. The perception of learning gains score from the researcher-developed instrument was calculated by subtracting student's self-reported score of what they believed they knew about skeletal muscle, pulmonary, and endocrine physiology at the beginning of the semester, from their self-reported score of how much they believed they learned by the end of the semester. The pre- and post-survey was administered using Qualtrics software (Version Jan., Feb. March, April, May, 2020; Qualtrics; Provo, UT, USA; https://www.qualtrics.com).

We measured student content knowledge after instruction in both treatment and comparison groups using a series of researcher-developed multiple-choice assessments, referred to as midterm content knowledge questions hereafter, that were included on each exam that aligned to specific course learning objectives in all sections. The midterm content knowledge questions were identical across three sections. None of the midterm content knowledge questions were the same as those used on the pre-survey, however, they tested the same concepts. We confirmed that the science concepts on the assessments were presented in each course and students were instructed to review the concepts before each exam. Human physiology and biology instruction experts evaluated the multiple-choice questions to ensure that the items assessed specific Human Anatomy and Physiology Society (HAPS) learning objectives and to categorize each item as either lower Bloom's learning levels (remember and understand) or higher Bloom's learning levels (apply, analyze, evaluate). The HAPS objectives assessed included G0306, G0402, G1301, M0102, M0203, M0301, M0305, M0307, M0602, M0603, M0608, M0614, M0704, M0801, M0901, J0301, J0302, J0401, J0402, J0406, J0501, J0503, J0901.

Instrument reliability

The instrument used to measure interest contained 11 items to measure maintained situational interest using a seven-point Likert scale (7 being strongly agree). According to Linnenbrink-Garcia et al. (2010), the Maintained-SI scale has good internal consistency, with a Cronbach alpha coefficient reported of 0.95 when item 9 and 11 were removed. In the current study, the Cronbach alpha coefficient was 0.94 for the post-survey when all 11 items were included.

The instrument used to measure expectancy, value, and cost contained ten items using a six-point Likert scale (6 being strongly agree). According to Getty et al. (2017) the EVC survey has good internal consistency among the individual components. In the current study, the Cronbach alpha coefficient was 0.88 for expectancy items, 0.82 for value items, and 0.80 for cost items, for the post-survey.

The instrument used to measure belonging to field and course contained five items each, using a seven-point Likert scale and had previously been used by Walton et al. (2015). In the current study, the Cronbach alpha coefficient was 0.79 for the belonging to field scale, and 0.82 for the belonging to course scale, for the post-survey.

Analysis approaches

Statistical analyses were performed using SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp.). Means are reported ± the standard deviation (*SD*). For all statistical tests, we used an alpha level of $p \leq 0.05$ for rejecting null hypotheses of equivalence of means.

To explore baseline equivalence for pre-survey scores between course sections (treatment, comparison 1, and comparison 2) for student learning gains, interest, motivation, and belonging to the field, we used a one-way between-groups analysis of variance (ANOVA). We conducted a one-way between-groups analysis of covariance (ANCOVA) to compare student learning gains and end of course interest, motivation, and belonging to the field among the three sections. The independent variable was the course section (treatment, comparison 1, and comparison 2), and the dependent variable consisted of the scores on the midterm content knowledge questions and post-survey (Figure 1). Students' scores on the pre-survey content knowledge questions, interest, motivation, and belonging to the field questions were used as the covariate. Where significant differences were found, we conducted pairwise comparisons to test for differences in the means among each group (treatment, comparison 1, and comparison 2).

Three outcome measures, students' sense of belonging to the course, students' perception of their learning gains and final grades only had post-course data. For these measures, we conducted a one-way between-groups analysis of variance (ANOVA) with post hoc comparisons using a Tukey Honestly Significant Difference (HSD) to test for differences in the means among each pairwise comparison.

We conducted a 2 by 2 between-groups analysis of covariance to explore the interaction of underrepresented race/ethnicity status, first generation status, and sex on student learning gains, interest, motivation, belonging to the field, between the treatment and comparison sections. Where significant interactions were found, we conducted a one-way ANCOVA to explore the various simple main

continued on next page

effects. Where significant differences in simple main effects were found, we conducted pairwise comparisons to test for differences in the means among groups.

A two-way between-groups analysis of variance was conducted to explore the impact of URM status, sex, and first gen status on student belonging to the course and final grades in lecture. Where significant interactions were found, we conducted a one-way ANOVA to explore the various simple main effects. Where significant differences in simple main effects were found, we conducted post hoc comparisons using a Tukey HSD to test for differences in the means among each pairwise comparison of groups.

Results

Baseline equivalence

There were no significant differences in the pre-survey scores between the treatment and comparison course sections for content knowledge (muscle: $F(2, 220) = 0.18$, $p = 0.84$, eta squared = 0.002, pulmonary: $F(2, 220) = 0.83$, $p = 0.44$, eta squared = 0.01, endocrine: $F(2, 220) = 1.37$, $p = 0.26$, eta squared =0.01), interest $F(2, 220) = 1.41$, $p = 0.25$, eta squared = 0.01, motivation (expectancy: $F(2, 220) = 0.26$, $p = 0.77$, eta squared = 0.002, value: $F(2, 220) = 0.86$, $p = 0.42$, eta squared =.01, cost: $F(2, 220) = 0.12$, $p = 0.89$, eta squared = 0.001), and belonging to field $F(2, 220) = 0.08$, $p = 0.93$, eta squared = 0.001 (Table 4).

The effect of using case studies during lecture, regardless of the delivery method (in person or online).

Student learning

After adjusting for pre-survey content knowledge scores, there was a significant difference between the mean treatment and comparison posttest scores for the content knowledge questions on exam 1: $F(2, 219) = 3.96$, $p = 0.02$, partial eta squared = 0.04, exam 2: $F(2, 218) = 9.13$, $p < 0.001$, partial eta squared = 0.08, and exam 3: $F(2, 216) = 48.325$, $p < 0.001$, partial eta squared = 0.31 (Figure 2). Pairwise comparisons of the adjusted means for exam 1 indicated that the treatment section scored significantly higher than the comparison 2 section ($p = 0.02$). There was not a significant difference between treatment and comparison 1 ($p = 0.14$) or comparison 1 and comparison 2 ($p = 1$) (Figure 2). Pairwise comparisons of the adjusted means for exam 2 indicated that the treatment section scored significantly higher than both comparison 1 ($p < 0.001$) and 2 ($p = 0.02$). There was not a significant difference between comparison 1 and comparison 2 ($p = 0.15$) (Figure 2). Pairwise comparisons of the adjusted means for exam 3 indicated that the treatment section scored significantly higher than both comparison 1 ($p < 0.001$) and 2 ($p < 0.001$). There was also a significant difference between comparison 1 and comparison 2 ($p < 0.001$), with comparison 2 scoring higher (Figure 2).

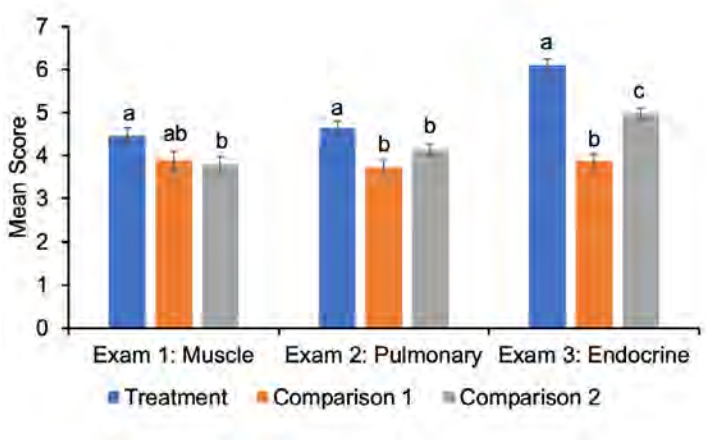

Figure 2. *Content knowledge scores. Average student score on posttest content knowledge questions after adjustment from pre-survey content knowledge scores. Scores were out of 7, 7, and 8 points for each exam respectively. Error bars represent ±SEM.*

Measure		Treatment (M±SD)	Comparison 1 (M±SD)	Comparison 2 (M±SD)
Content knowledge	Skeletal muscle	1.53 (0.9)	1.63 (1.0)	1.55 (1.1)
	Pulmonary	1.84 (1.1)	1.98 (1.2)	1.73 (1.1)
	Endocrine	1.84 (1.2)	1.67 (1.1)	1.56 (1.1)
Interest		63.91 (12.1)	66.87 (9.1)	63.95 (11.1)
Motivation	Expectancy	15.00 (2.7)	15.19 (2.3)	14.88 (2.4)
	Value	15.95 (2.0)	15.54 (2.4)	15.54 (2.3)
	Cost	12.43 (3.3)	12.33 (4.2)	12.63 (4.0)
Belonging to field		25.27 (5.5)	25.42 (5.1)	25.57 (4.8)

Table 4. *ANOVA baseline equivalence means and standard deviations*

continued on next page

Additionally, after adjusting for pre-survey content knowledge scores, there was a significant difference between the mean treatment, comparison 1, and comparison 2 posttest scores for the higher-level Bloom's content knowledge questions on all three exams combined $F(2, 219) = 4.78$, $p = 0.01$, partial eta squared = 0.04 with mean scores of 6.26 (*SD*=1.9), 5.25 (*SD*=1.9), and 5.64 (*SD*=1.9), for each section respectively, out of a possible 11 points. Pairwise comparisons of the adjusted means for the higher-level Bloom's content knowledge questions indicated that the treatment section scored significantly higher than the comparison 1 section ($p = 0.01$). There was not a significant difference between treatment and comparison 2 ($p = 0.10$) or comparison 1 and comparison 2 ($p = 0.73$).

Due to the unexpected transition to online learning caused by the COVID-19 pandemic, the University allowed students to choose a credit/no credit grading option. Students selecting this option were omitted from final grade analysis. One student from comparison 2 was omitted because data was unavailable for administrative reasons. After omitting students who opted for a credit/no credit grade, there was a statistically significant difference in students' final grade between the treatment and comparison sections $F(2, 212) = 16.87$, $p < 0.001$, eta squared = 0.14) (Figure 3). Post hoc comparisons using the Tukey HSD test indicated that the mean final grades for the treatment section were significantly higher than both comparison 1 ($p < 0.001$) and 2 sections ($p = 0.001$). Comparison sections 1 and 2 mean final grades also differed significantly ($p = 0.04$), with comparison 2 being higher (Figure 3).

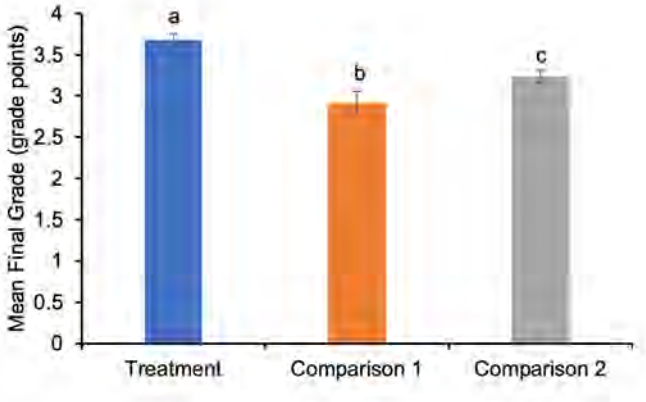

Figure 3. *Final grades. Average grades for students enrolled in the study in each of the three undergraduate human physiology sections offered during the spring semester 2020. Grades are reported using the 4-point scale. Error bars represent ±SEM.*

Student interest in the course

After adjusting for pre-survey scores, there was a significant difference between the mean treatment and comparison post-survey interest scores $F(2, 205) = 4.05$, $p = 0.02$, partial eta squared = 0.04 (Figure 4). Pairwise comparisons of the adjusted means for interest indicated that the treatment section scored significantly higher than the comparison 1 section ($p = 0.02$). There was not a statistical difference between treatment and comparison 2 ($p = 0.67$) or comparison 1 and comparison 2 ($p = 0.20$) (Figure 4).

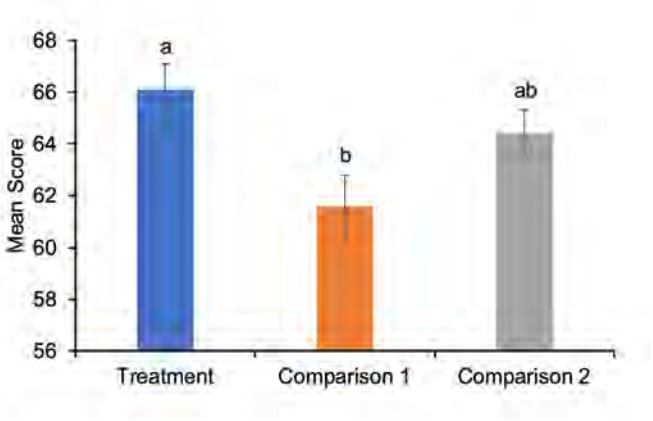

Figure 4. *Interest scores. Average student scores on post-survey interest questions after adjustment from pre-survey interest scores. Scores were out of a possible 77 points. Error bars represent ±SEM.*

Student motivation in the course

After adjusting for pre-survey expectancy, value, and cost scores, there was a significant difference between the mean treatment and comparison posttest scores for expectancy $F(2, 205) = 5.79$, $p = 0.004$, partial eta squared=0.05 and cost $F(2, 205) = 3.84$, $p = 0.02$, partial eta squared = 0.04. There was no significant difference between treatment and comparison groups for value $F(2, 205) = 1.72$, $p = 0.18$, partial eta squared = 0.02 (Figure 5). Pairwise comparisons of the adjusted means for expectancy indicated that the treatment section scored significantly higher than both the comparison 1 ($p = 0.01$) and comparison 2 sections ($p = 0.02$). There was not a significant difference between comparison 1 and comparison 2 ($p = 1$) (Figure 5). Pairwise comparisons of the adjusted means for cost indicated that the treatment section scored significantly lower

continued on next page

than comparison 1 ($p = 0.02$). There was not a significant difference between treatment and comparison 2 ($p = 0.33$) or comparison 1 and comparison 2 ($p = 0.48$) (Figure 5).

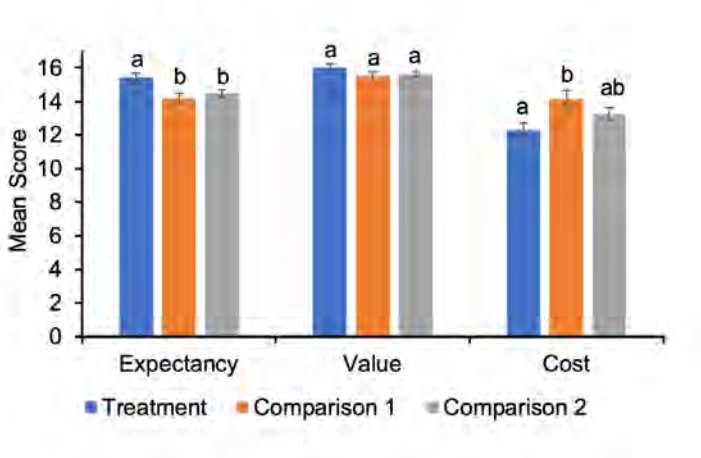

Figure 5. *Motivation scores. Average student scores on post-survey expectancy, value, and cost questions after adjustment from pre-survey expectancy, value, and cost scores. Scores were out of 18, 18, and 24 points respectively. Error bars represent ±SEM.*

Student's sense of belonging

There was a statistically significant difference in the post-survey belonging to course scores between the treatment and comparison course sections $F(2, 206) = 8.52$, $p < 0.001$, eta squared = 0.08 (Figure 6). Post hoc comparisons using the Tukey HSD test indicated that the treatment group scored significantly higher than both comparison section 1 ($p = 0.001$) and 2 ($p = 0.002$). There was not a significant difference between comparison 1 and comparison 2 sections ($p = 0.82$) (Figure 6).

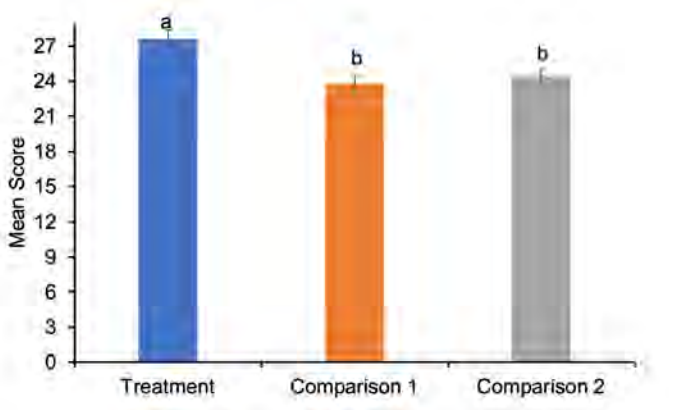

Figure 6. *Belonging to course scores. Average student scores on post-survey belonging to course questions. Scores were out of 35 points. Error bars represent ±SEM.*

After adjusting for pre-survey scores, there was not a significant difference between treatment and comparison sections on the post-survey belonging to field score $F(2, 205) = 1.80$, $p = 0.17$, partial eta squared = 0.02 with adjusted mean scores of 26.1 ($SD = 4.3$) for the treatment section, 24.67 ($SD = 4.3$) for comparison 1, and 25.14 ($SD = 4.3$) for comparison 2.

Student's assessment of their learning gains

After calculating a mean gains score for students' perception of how much they understood each topic at the beginning of the semester versus the end, there was no significant difference between treatment and comparison sections for the muscle topic $F(2, 206) = 0.32$, $p = 0.72$, eta squared = 0.003, the pulmonary topic $F(2, 206) = 0.67$, $p = 0.52$, eta squared = 0.01, or the endocrine topic $F(2, 206) = 1.36$, $p = 0.26$, eta squared= 0.01).

As a part of the modified SALG on the post-survey we asked students which aspects of the class helped their learning and helped them connect scientific concepts to their everyday life. Students in the treatment section believed the case studies helped a moderate to good amount. Additionally, when comparing the three treatment groups, students in the treatment section believed that instructor lectures and class discussions were more helpful (Tables 5 and 6).

Item	Treatment (M±SD)	Comparison 1 (M±SD)	Comparison 2 (M±SD)
Attending or listening to instructor's lectures	5.27 (0.9)	4.63 (1.4)	4.62 (1.2)
Completing case studies	4.90 (1.1)	—[a]	—[a]
Participating in discussions during class	4.31 (1.5)	3.76 (1.5)	3.23 (2.0)
Reading the textbook	3.59 (1.6)	3.00 (1.5)	3.95 (1.5)
Note. Likert scale key is as follows: 1-NA, 2-provided no help, 3-helped a small amount, 4-helped a moderate amount, 5-helped a good amount, 6-helped a great amount. Treatment n=71, comparison 1 n=51, and comparison 2 n=87.			
[a] Means and SD for this item was excluded since comparison groups did not receive case-based instruction.			

Table 5. *Means and standard deviations for the question: Overall, how much did each of the following aspects of the class help your learning?*

continued on next page

Item	Treatment (M±SD)	Comparison 1 (M±SD)	Comparison 2 (M±SD)
Attending or listening to instructor's lectures	5.15 (1.0)	4.65 (1.4)	4.74 (1.1)
Completing case studies	5.06 (1.1)	—[a]	—[a]
Participating in discussions during class	4.35 (1.5)	3.96 (1.5)	3.53 (1.9)
Reading the textbook	3.45 (1.6)	3.16 (1.6)	3.95 (1.6)
Note. Likert scale key is as follows: 1-NA, 2-provided no help, 3-helped a small amount, 4-helped a moderate amount, 5-helped a good amount, 6-helped a great amount. Treatment n=71, comparison 1 n=51, and comparison 2 n=87.			
[a] Means and SD for this item was excluded since comparison groups did not receive case-based instruction.			

Table 6. *Means and standard deviations for the question: Overall, how much did each of the following aspects of the class help you to understand the connections between scientific concepts and other aspects of your everyday life?*

Sex, URM, and first-generation status

Analyses of interactions among moderating demographic variables showed that most interactions were not significant. The following results are the few significant interactions and simple main effects we found.

A generation (first vs non first) x treatment groups (treatment, comparison 1, comparison 2) ANCOVA was conducted on exam 1, controlling for pre-survey content knowledge scores. Results found a significant main effect for treatment groups $F(2, 181) = 4.73$, $p = 0.01$, partial eta squared = 0.05, and no significant main effect for generation status $F(1, 181) = 0.95$, $p = 0.33$, partial eta squared = 0.01. Results also found a significant interaction $F(2, 181) = 4.17$, $p = 0.02$, partial eta squared = 0.04.

Simple main effect follow-up analyses for first generation college students found that there was a significant difference between the mean treatment and comparison posttest content knowledge scores for first generation students taking exam 1 $F(2, 100) = 9.57$, $p < 0.001$, partial eta squared = 0.16) (Table 7). Pairwise comparisons indicated that the treatment section scored significantly higher than both the comparison 1 ($p < 0.001$) and comparison 2 sections ($p = 0.002$). There was no significant difference between comparison 1 and comparison 2 ($p = 0.95$) (Table 7).

For non-first generation college students, no significant differences were found among treatment groups $F(2, 80) = 0.59$, $p = 0.56$, partial eta squared = 0.02.

Treatment Groups	Adjusted Mean	SD	n	F	partial η^2
Treatment	4.92 a	1.4	32	9.57*	0.16
Comparison 1	3.36 b	1.6	28		
Comparison 2	3.72 b	1.5	44		
Note. *$p < 0.001$. Score for exam 1 was out of 7.					
Means with different letters next to them denote significant differences.					

Table 7. *ANCOVA results comparing treatment groups on exam 1 for first generation college students*

A sex (male vs female) x treatment groups (treatment, comparison 1, comparison 2) ANCOVA was conducted on exam 3, controlling for pre-survey content knowledge scores. Results found a significant main effect for treatment groups $F(2, 212) = 37.2$, $p < 0.001$, partial eta squared = 0.26, and no significant main effect for sex $F(1, 212) = 1.36$, $p = 0.25$, partial eta squared = 0.01. Results also found a significant interaction $F(2, 212) = 3.26$, $p = 0.04$, partial eta squared = 0.03.

Simple main effect follow-up analyses for female college students found that there was a significant difference between the mean treatment and comparison posttest content knowledge scores for female students taking exam 3 $F(2, 143) = 44.93$, $p < 0.001$, partial eta squared = 0.39 (Table 8). Pairwise comparisons indicated that the treatment section scored significantly higher than both the comparison 1 ($p < 0.001$) and comparison 2 sections ($p < 0.001$). There was also a significant difference between comparison 1 and comparison 2 ($p < 0.001$), with comparison 2 scoring higher (Table 8).

Treatment Groups	Adjusted Mean	SD	n	F	partial η^2
Treatment	6.24 a	1.3	51	44.93*	0.39
Comparison 1	3.69 b	1.3	34		
Comparison 2	4.81 c	1.2	62		
Note. *$p < 0.001$. Score for exam 3 was out of 8.					
Means with different letters next to them denote significant differences.					

Table 8. *ANCOVA results comparing treatment groups on exam 3 for female students*

continued on next page

For male students taking exam 3, simple main effect follow up analyses found that there was a significant difference between the mean treatment and comparison posttest content knowledge scores $F(2, 68) = 8.24$, $p = 0.001$, partial eta squared = 0.20 (Table 9). Pairwise comparisons indicated that the treatment section scored significantly higher than the comparison 1 section ($p < 0.001$), but not the comparison 2 section ($p = 0.28$). There was also a significant difference between comparison 1 and comparison 2 ($p = 0.03$), with comparison 2 scoring higher (Table 9).

Treatment Groups	Adjusted Mean	SD	n	F	partial η^2
Treatment	5.83 a	1.2	25	8.24*	0.2
Comparison 1	4.24 b	1.2	17		
Comparison 2	5.24 a	1.3	30		
Note. *$p < 0.05$. Score for exam 3 was out of 8.					
Means with different letters next to them denote significant differences.					

Table 9. *ANCOVA results comparing treatment groups on exam 3 for male students*

A sex (male vs female) x treatment groups (treatment, comparison 1, comparison 2) ANOVA was conducted on final grades in the course. Results found a significant main effect for treatment groups $F(2, 209) = 11.08$, $p < 0.001$, partial eta squared = 0.10, and no significant main effect for sex $F(1, 209) = 0.53$, $p = 0.47$, partial eta squared = 0.003. Results also found a significant interaction $F(2, 209) = 3.30$, $p = 0.04$, partial eta squared = 0.03.

Simple main effect follow-up analyses for female college students found that there was a significant difference between the mean treatment and comparison sections' final grades $F(2, 141) = 19.43$, $p < 0.001$, eta squared = 0.22 (Table 10). Post hoc comparisons using the Tukey HSD test indicated that females in the treatment group received significantly higher final grades in the course than females in both comparison group 1 ($p < 0.001$) and 2 ($p < 0.001$). There was also a significant difference between females in comparison 1 and comparison 2 sections ($p = 0.01$), with females in comparison 2 receiving higher grades (Table 10).

For male college students, no significant differences were found among treatment groups $F(2, 68) = 0.94$, $p = 0.40$, eta squared = 0.03.

Treatment Groups	Mean	SD	n	F	η^2
Treatment	3.78 a	0.5	50	3.30*	0.22
Comparison 1	2.76 b	1.1	33		
Comparison 2	3.22 c	0.7	61		
Note. *$p < 0.001$. Grades are reported using the 4-point scale.					
Means with different letters next to them denote significant differences.					

Table 10. *ANOVA results comparing treatment groups on final grades for female students*

Discussion

The purpose of this quasi-experiment was to test the hypothesis that integrating case studies into an introductory human physiology course would increase students' learning gains, interest, motivation, and sense of belonging compared to the two comparison sections that did not use case studies.

Exploring outcomes in student learning

Our results show that on all three exams, the treatment group scored significantly higher on the post-survey content knowledge questions than one or in some cases both of the comparison sections. Exam 1 was the only exam given in an in-person classroom setting, and was the only exam in which the treatment section scored significantly higher than only one of the comparison sections, comparison 2. While not significant, the difference in means between the treatment and comparison 1 section was nearly the same as the difference in means between the treatment and comparison 2, but the smaller sample size in comparison 1 led to lower statistical power. Exams 2 and 3 were given online after the switch to virtual learning caused by the COVID-19 pandemic, and in both cases the treatment section scored significantly higher than both of the comparison sections. Exam 3 which covered the endocrine system had the largest difference with the treatment section scoring 58% higher than comparison 1 and 22% higher than comparison 2. Our results are similar to the findings of Cliff and Wright (1996) who found that the addition of case studies to the anatomy and physiology course resulted in an increase in exam scores and the general findings for positive impacts of active learning (Freeman et al. 2014). The general increase in improvement of the scores of students learning with case studies may be explained by differences in the cases themselves, the benefits of continued practice

continued on next page

with case studies, or that conducting case studies online may be more beneficial. Future studies that explore the impact of case study dosage would be valuable, as would studies that rigorously explore using case studies in online versus face-to-face settings.

Case studies have been hypothesized to increase students' ability to think critically, improving problem solving skills and depth of learning (Allchin 2013; Cliff and Wright 1996; Herreid 1994; Herreid 2004; Murray-Nseula 2011; Smee and Cooke 2018; Tomey 2003; Willcox 1999; Yadav et al. 2007). When our post-survey content knowledge questions on the exams were separated by Blooms level and analyzed, we found that the treatment section scored significantly higher than the comparison 1 section for the higher-level Blooms questions. This is significant because students being able to answer higher level Blooms questions conveys a deeper understanding of the content and ability to think critically which is important for students in the health science field.

Our results also show that students in the treatment section received significantly higher final grades than both comparison sections which has important implications for student persistence in the health science professions and for possibly decreasing failure or repeat rates. These results are compelling and are suggestive that some factor in the treatment section, namely the use of case studies, has led to the higher overall performance of the treatment section compared to the comparison sections, though the quasi-experimental design of the study does not allow us to reject alternative explanations for these differences. The results of analyses of data disaggregated by demographic groups suggest that the improvement in grades for the case studies class was driven primarily by increases in the success of students who identify as women. We encourage additional replications of this type of research on case studies to see if this result applies more generally.

The unique situation created by the response to COVID-19 where all learning was transitioned to an online format midsemester is not lost on us. The continued out-performance of the treatment group compared to the comparison groups despite the transition midsemester may suggest that case studies are effective for use in both in-person and online learning formats.

Exploring outcomes in student interest

Student interest in physiology is an important outcome in and of itself and may significantly affect other academic outcomes such as achievement (National Academies of Sciences 2017). Our results demonstrate that the treatment section had higher levels of interest than our comparison 1 section. Though our treatment section scored higher on interest than our comparison 2 section, the difference was not statistically significant. Effect sizes were small to medium. It is interesting to note that the comparison 1 section had lower final grades when compared to the treatment and comparison 2 section. It may be possible that this low level of interest expressed in the comparison 2 section contributed negatively to final grades in the course while the higher level of interest in the treatment section contributed positively to overall grades in the course. We cannot be sure that the case studies in the treatment section are responsible for the higher level of interest expressed.

Exploring outcomes in student motivation

The instrument we used for a motivation construct includes three factors that contribute to motivation: expectancy, value, and cost. Creators of the measure highlighted its usefulness for measuring the effects interventions may have on student motivation and its ability to predict how these three factors affect student achievement and interest (Getty et al. 2017). Although students in all three courses started with similar levels of expectancy, value, and cost, our results demonstrate that students in the treatment section had higher levels of expectancy at the end of the course, compared to our comparison sections. Students in the treatment section also had lower levels of cost compared to the comparison 1 section, meaning that students in the treatment section reported fewer barriers to being successful in the course, such as time and effort. Again, though our treatment section had lower levels of cost than our comparison 2 section, the difference was not statistically significant. There was no difference between treatment groups for value. This finding was surprising to us and requires further investigation.

Exploring outcomes in student sense of belonging

The National Academies of Science have identified belonging as an important core competency related to student achievement and indicated that its development may be most beneficial for students from underrepresented groups (National Academies of Sciences 2017). We used two different measures to evaluate student sense of belonging. The first measure asked about students' sense of belonging to the health science field. Baseline equivalency was established amongst all three sections for this measure. Our results showed that there was no difference among treatment groups for students' sense of belonging to the field at the end of the course. Student sense of belonging actually remained relatively unchanged from the beginning of the semester to the end across all three treatment groups. This was somewhat surprising to us since we had hypothesized that the case studies in the treatment group would increase students' sense of belonging to the field, yet in other respects it was not too surprising because only three case studies were used in this quasi-experiment.

The second measure we used asked about students' sense of belonging in the course itself which was only included on the post-survey. Our results showed that students in the treatment section felt as though they belonged to the course significantly more than the comparison sections with moderate effect sizes. We suspect that the use of case studies and the format of students working together on them may have contributed to the increased sense of belonging in the

continued on next page

treatment section. With the transition to online learning mid semester, this result of increased belonging in the treatment section is particularly noteworthy and again speaks to the value of using case studies.

Exploring outcomes in student assessment of learning gains

A measure was included on the post-survey asking students about how much they perceived they learned about each of the researched systems (skeletal muscle, pulmonary, and endocrine). There was no difference found between treatment groups. These results are somewhat similar to those from a study by Deslauriers et al. (2019), which found that students in active learning classes in physics feel like they learned less than their peers in passive learning classes even though they actually learned more.

When asked how well they felt the case studies helped with their learning and ability to make connections between what they learned and their everyday life, students in the treatment section, on average, believed the case studies helped "a good amount". This is similar to what Bonney (2015) found using a similar modified SALG instrument. Wilcox (1999) gave students a Likert scale survey asking students about the usefulness of case studies in the course in which most students responded that they found the case studies "useful".

Exploring outcomes for disaggregated demographic groups

When comparing outcomes for various demographic groups (male, female, URM, non-URM, first gen, and non-first gen), our results demonstrated that very few differences were found between treatment groups. When differences were found, they tended to favor underrepresented groups in the treatment section. The most notable difference was found in the results of final grades for females described earlier. Other differences found include first generation students in the treatment section scoring significantly higher than comparison sections on the content knowledge questions for exam 1, and female students in the treatment section scoring significantly higher than comparison section for the content knowledge questions on exam 3.

Conclusion

Using case studies to help teach human physiology is a worthwhile endeavor. Students generally feel like they are helpful and the use of just three case studies throughout a semester resulted in positive marked differences in exam scores, final grades, ability to learn and understand concepts, interest, expectancy to succeed, lower costs, and an improved sense of belonging to the course which may be attributed to the case studies. Further studies using more rigorous experimental designs are warranted to further analyze the impact of case studies on student learning and motivation.

Limitations

As any study that involves human subjects, our study does have several limitations. One factor that may explain the differences we observed are instructor related. As a quasi-experiment, we did not control for differences in instructor experience, personality, or even teaching style which is why we called our non-treatment groups our "comparison" groups rather than "control" groups. Additionally, with only one treatment section it is possible that an instructor-related variable outside the use of case studies could explain the differences observed between sections. We were limited to only having three sections of the course available for analysis and chose to implement case studies in one because it better enabled us to explore possible confounding factors that were distributed among the two comparison groups. For example, class size varied among the three sections (from 60-101 students). However, the influence of class size on student achievement is questionable for courses larger than 30 (Ake-Little 2020) and the treatment section class size was intermediate between the two comparison sections. Other possible confounding factors that we were able to consider included the use of learning assistants by both the treatment and comparison 1 sections, and the number of times a week a class was held prior to moving online. This study attempted to balance instructor differences by assigning the least experienced instructor to teach the treatment section.

A general critique of quasi-experiments is that student enrollment in different sections is not random. To help address this concern, we conducted baseline comparisons. All three course sections showed baseline equivalence for pre-survey content knowledge, interest, motivation, and belonging to the field. This means that each section can be assumed to have started the course at the same level of understanding, interest, motivation, and belonging, increasing the likelihood that we can attribute changes in outcome variables to differences in pedagogy.

Another limitation to the study is student attendance. Due to the large class size in each course section, attendance was not taken and it is quite possible that student absences may have contributed negatively towards grades. This problem may have been more significant if students missed the day the case studies were completed in class or did not complete the assignment at all.

Mid semester, the COVID-19 pandemic necessitated a transition to online learning. In many ways the transition to online teaching made the classes more equivalent. Prior to the transition to online learning, one of the comparison sections was scheduled to hold class three days a week while the other comparison section and the treatment section were scheduled to hold class twice a week. With the transition to online learning, all instructors delivered their course content in the same manner, asynchronously once a week. The transition to online learning warrants future work. In this study, one case study was implemented before the transition and the results for the impact of the case study on student achievement were consistent (though the differences in means were slightly lower) with results from case studies conducted after the transition to virtual teaching. However, because of the difference in the way the first case study was conducted (in

continued on next page

person), versus the second and third case studies (online), care must be taken when comparing the results to each other. For example, exam 1 was proctored in person, meaning that there were fewer opportunities for students to cheat compared to the online format for exams 2 and 3. While there is no evidence of significant cheating on exams, the possibility remains. The transition to remote learning during this semester also brought with it profound psychological effects due to fear, mandates on social distancing, and inadequate home learning environments to name a few. The transition to remote learning compounded with these psychological effects may have had an impact on students' interest, motivation and sense of belonging.

Acknowledgments

We thank the lecturers for allowing us to conduct research in their courses. Without their unwavering support, especially in light of having to transition to online learning, this study would not have been possible. We would also like to thank the CPP Registrar Office for providing us with University data, the many students who participated in the study, and the NCCSTS for providing a valuable platform for case studies.

About the Authors

All the authors are affiliated with California State Polytechnic University at Pomona. Mackenzie Pekary, PhD, is a recent graduate who teaches Human Physiology laboratory. Paul Beardsley, PhD, is a professor in the Center for Excellence for Mathematics and Science Teaching and the Biological Sciences Department. Juanita Jellyman, PhD, is an associate professor in the Biological Sciences Department who teaches Human Physiology and Endocrinology. Michael Giang, PhD, is an associate professor in the Psychology Department who teaches research design courses.

References

Association of American Medical Colleges. 2018. Percentage of all active physicians by race/ethnicity. [cited 2020 Oct 30]. Available from: https://www.aamc.org/data-reports/workforce/interactive-data/figure-18-percentage-all-active-physicians-race/ethnicity-2018.

Ake-Little E, von der Embse N, Dawson D. 2020. Does class size matter in the university setting? *Educ Res* 49(8):595–605. https://doi.org/10.3102/0013189X20933836.

Allchin D. 2013. Problem- and case-based learning in science: An introduction to distinctions, values, and outcomes. *CBE Life Sci Educ* 12(3):364–372. https://doi.org/10.1187/cbe.12-11-0190.

Alsan M, Garrick O, Graziani GC. 2019. Does diversity matter for health? Experimental evidence from Oakland. *Am Econ Rev* 109(12):4071-4111. https://doi.org/10.1257/aer.20181446

Barron KE, Hulleman CS, Getty SG, Taylor J. 2017. Expectancy value-cost survey of student motivation: User guide. [internet] Charlottesville, VA: The Motivate Lab. Available from: https://www.researchgate.net/publication/319433479_User's_Guide_for_the_Expectancy-Value-Cost_Survey_of_Student_Motivation

Birk T, Lee V, Corbit A. 2019. The effects of sex, ethnicity, and socioeconomic status on student perceptions of case-based learning in anatomy and physiology classes. *HAPS Educator* 23(2):385-395. https://doi.org/10.21692/haps.2019.022.

Bonney KM. 2015. Case study teaching method improves student performance and perceptions of learning gains. *J Microbiol Biol Educ* 16(1):21–28. https://doi.org/10.1128/jmbe.v16i1.846.

Canning EA, Harackiewicz JM, Priniski SJ, Hecht CA, Tibbetts Y, Hyde JS. 2018. Improving performance and retention in introductory biology with a utility-value intervention. *J Educ Psychol* 110(6):834–849. https://doi.org/10.1037/edu0000244.

Cheruvelil KS, Soranno PA, Weathers KC, Hanson PC, Goring SJ, Filstrup CT, Read EK. 2014. Creating and maintaining high-performing collaborative research teams: the importance of diversity and interpersonal skills. *Front Ecol Environ* 12(1):31–38. https://doi.org/10.1890/130001.

Cliff WH, Wright AW. 1996. Directed case study method for teaching human anatomy and physiology. *Adv Physiol Educ* 270(6): S19. https://doi.org/10.1152/advances.1996.270.6.S19

Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. 2019. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *PNAS* 116(39):19251–19257. https://doi.org/10.1073/pnas.1821936116.

Eccles J. 1983. Expectancies, values, and academic behaviors. San Francisco (CA): WH Freeman. p. 7-68.

Ediger TL. 2017. What, how, and why: Writing interview-based cases helps students connect anatomy and physiology concepts to real life. *HAPS Educator* 21(2):148–151. https://doi.org/10.21692/haps.2017.031.

Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. *PNAS* 111(23):8410–8415. https://doi.org/10.1073/pnas.1319030111.

continued on next page

Getty SG, Hulleman CS, Barron KE, Ruzek ER, Flake JK, Foley K. 2017. Using the expectancy-value-cost theory of motivation to understand student achievement and future interest in STEM classrooms [White paper]. Charlottesville (VA): The Motivate Lab [cited 2020 Oct 30]. Available from: https://www.researchgate.net/profile/Chris_Hulleman/publication/326354387_A_Rapid_Measure_of_Student_Motivation-Using_an_Expectancy-Value-Cost_Model_of_Motivation_to_Understand_Student_Achievement_and_Future_Interest_in_STEM_Classrooms/links/5b47803c0f7e9b4637d1b979/A-Rapid-Measure-of-Student-Motivation-Using-an-Expectancy-Value-Cost-Model-of-Motivation-to-Understand-Student-Achievement-and-Future-Interest-in-STEM-Classrooms.pdf

Hidi S, Renninger KA. 2006. The four-phase model of interest development. *Educ Psychol* 41(2):111–127. https://doi.org/10.1207/s15326985ep4102_4.

Herreid CF. 1994. Case studies in science--A novel method of science education. *J Coll Sci Teach* 23(4):221–29.

Herreid CF. 1997. Bringing to science education the established teaching tool of law and medicine. *J Coll Sci Teach* 27(2): 92-94.

Herreid CF. 2004. Can case studies be used to teach critical thinking? *J Coll Sci Teach* 33(6):12-14.

Herreid CF. 2005. Using case studies to teach science. *American Institute of Biological Sciences*.

Leavitt HL. 2018. Asthma Attack! Buffalo (NY): National Center for Case Study Teaching in Science.

Linnenbrink-Garcia L, Durik AM, Conley AM, Barron KE, Tauer JM, Karabenick SA, Harackiewicz JM. 2010. Measuring situational interest in academic domains. *Educ Psychol Meas* 70(4):647–671. https://doi.org/10.1177/0013164409355699.

Murray-Nseula M. 2011. Incorporating case studies into an undergraduate genetics course. *Journal of the Scholarship of Teaching and Learning* 11(3):75–85.

National Academies of Sciences, Engineering, and Medicine. 2017. Supporting students' college success: The role of assessment of intrapersonal and interpersonal competencies. Washington, DC: National Academies Press. https://doi.org/10.17226/24697.

Neumann RT, Quinn CJ, Whitaker BA, Woyton ST, Harris BN. 2016. All or nothing: a case study in muscle contraction. Buffalo (NY): National Center for Case Study Teaching in Science.

Schillo KK. 2012. Muscleman: a surprising case of shrinkage. Buffalo (NY): National Center for Case Study Teaching in Science.

Smee D, Cooke J. 2018. Making it real: Case-study exam model. *HAPS Educator* 22(3):268–271. https://doi.org/10.21692/haps.2018.029.

Theobald EJ, Hill MJ, Tran E, Agrawal S, Arroyo EN, Behling S, Chambwe N, Cintrón DL, Cooper JD, Dunster G, Grummer JA, Hennessey K, Hsiao J, Iranon N, Jones II L, Jordt H, Keller M, Lacey ME, Littlefield CE, Lowe A, Newman S, Okolo V, Olroyd S, Peecook BR, Pickett SB, Slager DL, Caviedes-Solis IW, Stanchak KE, Sundaravardan V, Valdebenito C, Williams CR, Zinsli K, Freeman S. 2020. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *PNAS* 117(12):6476–6483. https://doi.org/10.1073/pnas.1916903117.

Tomey AM. 2003. Learning with cases. *J Contin Educ Nurs* 34(1):34–38. https://doi.org/10.3928/0022-0124-20030101-07.

Walton GM, Logel C, Peach JM, Spencer SJ, Zanna MP. 2015. Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *J Educ Psych* 107(2):468–485. https://doi.org/10.1037/a0037461.

Wilcox KJ. 1999. The case method in introductory anatomy & physiology: Using the news. *Am Biol Teach* 61(9):668–671. https://doi.org/10.2307/4450799.

Yadav A, Lundeberg M, DeSchryver M, Dirkin K, Schiller NA, Maier K, Herreid CF. 2007. Teaching science with case studies: A national survey of faculty perceptions of the benefits and challenges of using cases. *J Coll Sci Teach* 37(1): 34-38.

Back to TOC