Impact of Supplemental Instruction Frequency and Format on Exam Outcomes

What the Flip? A Pilot Study

Ten Years in the A&P1 Classroom - A Retrospective Analysis

Confidential Peer-Evaluation and Online Learning

Online Tutoring System (MoFaCTS) for Anatomy & Physiology
At Northeast College of Health Sciences, you can combine your content expertise with pedagogy development to become an in-demand anatomy and physiology instructor. Our program gives you the advantages of:

- A 100% online, asynchronous program, for ultimate flexibility.
- Study with leading practitioners, including the past president of the Human Anatomy and Physiology Society (HAPS), and three recipients of HAPS' President's Medal Award.
- Small class sizes, to enhance learning.

You’ll graduate with a deep understanding of science-based techniques, pedagogy, and human structure and function, ready to succeed as a virtual instructor or advance your career in traditional classroom, online, and hybrid settings.

To learn more, visit northeastcollege.edu/HAPI.

“Even the most experienced clinicians and educators often find that, despite their subject matter expertise, making the transition to virtual instructor can be fraught with difficulties if not preceded by specific training.”

- Dr. Peter Nickless, Northeast College Dean of Online Education
TABLE OF CONTENTS

EDUCATIONAL RESEARCH

Impact of Supplemental Instruction Frequency and Format on Exam Performance in Anatomy and Physiology.
https://doi.org/10.21692/haps.2022.013
Blase Rokusek, BS, Emilee Moore, BS, Christopher Waples, PhD, Janet Steele, PhD .............................................................. 5

PERSPECTIVES ON TEACHING

What the Flip? A Pilot Study Perspective on the Flipped Classroom for Medical School Physiology.
https://doi.org/10.21692/haps.2022.006
Chasity B. O’Malley, PhD ....................................................................................................................................................................................... 14

Ten years in the Human Anatomy and Physiology I Classroom: A Retrospective Analysis of Student Preparation,
https://doi.org/10.21692/haps.2022.010
Carol Britson, PhD .................................................................................................................................................................................................... 19

Confidential Peer-Evaluation as a Method of Learning in Online University Courses.
https://doi.org/10.21692/haps.2022.011
Simon Lemaire, PhD, Gladys Bruyninx, MD, Miriam Grenon, CL, Madisson Kellerher-Radey, BScN,
Alexander Yeuchyk, PhD ....................................................................................................................................................................................... 37

Online Tutoring System (MoFaCTS) for Anatomy and Physiology: Implementation and Initial Impressions.
https://doi.org/10.21692/haps.2022.012
Amanda M. Banker, MS, Philip I. Pavlik Jr, PhD, Andrew Olney, PhD, Luke G. Eglington, PhD .................................................. 44

HAPS Committees and Boards ...................................................................................................................................................................................... 55
**HAPS Educator Journal of the Human Anatomy and Physiology Society**

**Editor-in-Chief** – Jackie Carnegie  
**Managing Editor** – Brenda del Moral

---

**Editorial Board**

Kerry Hull  
Carol Britson  
Jackie Carnegie  
Greg Crowther  
Brenda del Moral  
Tracy Ediger  
Elizabeth Granier  
Justin Shaffer  
Zoe Soon

---

**Reviewer Panel**

Jessica Adams  
Teresa Alvarez  
Aakanksha Angra  
Jordan Balaban  
Freddie Bauer  
Amy Bauguess  
Kathy Bentons  
Carol Britson  
Kathy Burleson  
Jackie Carnegie  
Janet Casagrand  
Keely Cassidy  
James Clark  
Anne Clayton  
Greg Crowther  
Leslie Day  
Brenda del Moral  
Jennifer Dennis  
Michel Desilets  
James Doyle  
Tracy Ediger  
Hisham Elbatarny  
David Evans  
Jun Farnsworth  
Matthew Fisher  
Camille Freeman  
Mindi Fried  
Nataliya Galifianakis  
Rhonda Gamble  
Anya Goldina  
Elizabeth Granier  
Jennifer Hillier  
Rose Hyson  
Michael Ibiwoye  
Jon Jackson  
Adrienne King  
Sara Klender  
Barbie Klein  
Rowena Korpal  
Tres Kuchter  
Richelle Laipply  
Alice Lawrence  
Paul Lyster  
Laylonda Maines  
Robert McCarthy  
Roberta Meehan  
Amber Miller  
Michele Moore  
Tracy Mowery  
Soma Mukhopadhyay  
Chasity O’Malley  
Zvi Ostrin  
Ana Maria Oyarce  
Andrew Petto  
Kristen Platt  
Katrina Porter  
Disa Pryor  
Melissa Quinn  
Wendy Rappazzo  
Ed Raynes  
Peter Reuter  
Wendy Riggs  
Rebecca Romine  
Krista Rompolski  
Heather Rudolph  
Usha Sankar  
Stephen Sarikas  
Joanne Savory  
Heidi Schutz  
Derek Scott  
Justin Shaffer  
Disa Smee  
Lola Smith  
Zoe Soon  
Maria Squire  
Leslie Stone-Roy  
Parker Stuart  
Sheela Venu  
Dani Waters  
Michael Waterson  
Adrienne Williams  
Jonathan Wisco

---

**Digital Media UX** – L. Katie Roberts

The HAPS Educator, The Journal of the Human Anatomy and Physiology Society, aims to foster teaching excellence and pedagogical research in anatomy and physiology education. The journal publishes articles under three categories. Educational Research articles discuss pedagogical research projects supported by robust data. Perspectives on Teaching articles discuss a teaching philosophy or modality but do not require supporting data. Current Topics articles provide a state-of-the-art summary of a trending topic area relevant to anatomy and physiology educators. All submitted articles undergo peer review. Educational Research articles will additionally be reviewed for the quality of the supporting data. All issues of the HAPS Educator are freely available, and individual articles are uploaded to the Life Science Teaching Resource Community (and link to https://www.lifescitrc.org/) and available in the Education Resources Information Center (ERIC).

The HAPS Educator is published in April, August and December. The deadlines for submission are March 1, July 1 and November 1.

Submission Guidelines for Authors

Information for authors on the terms of submission, the submission procedure, formatting the manuscript, formatting the references, the submission of illustrations, and the peer review process, is available HERE.

Submission Link

When ready, please follow this link to submit your manuscript to the HAPS Educator.

---

You do not need to be a member of the Human Anatomy and Physiology Society (HAPS) to publish in the HAPS Educator. For more information see the complete submission guidelines using the link above.

---

**Human and animal research subjects**

Research that includes dissection and manipulation of animal tissues and organs must adhere to the Human Anatomy and Physiology Society (HAPS) Position Statement on Animal Use, which states that the use of biological specimens must be in strict compliance with federal legislation and the guidelines of the National Institutes of Health and the United States Department of Agriculture.

The use of humans or animals in research must fulfill clearly defined educational objectives.

Experimental animals must be handled in accordance with the author’s institutional guidelines and informed consent must be obtained for studies on humans. It is the responsibility of the author(s) to secure IRB approval for research on humans.

**Plagiarism**

Authors must obtain permission to reproduce any copyright material and the source of this material must be acknowledged in their manuscript.

**Disclaimer**

Responsibility for (1) the accuracy of facts, (2) the expression of opinion and (3) the authenticity of any supporting material presented by the author rests solely with the author. The HAPS Educator, its publishers, editors, reviewers and staff, take no responsibility for these things.

**CONTACT THE HAPS-Educator Editor** if you have additional questions or concerns.

The HAPS Educator is published electronically by The Human Anatomy and Physiology Society (HAPS). The written and visual contents of this magazine are protected by copyright. Temporary permission is granted for members of the Human Anatomy and Physiology Society to read it on-line, to print out single copies of it, and to use it unchanged for any non-commercial research and educational purpose, including making copies for classroom use provided the materials are not modified and appropriate acknowledgment is made of the source. All other uses of this material are conditional and require the consent of the editor - and when applicable, the other copyright owners. Requests for permission should be directed to the editor via the contact information stated above.

© August, 2022 All rights reserved.
Impact of Supplemental Instruction Frequency and Format on Exam Performance in Anatomy and Physiology

Blase Rokusek, BS¹⁺², Emilee Moore, BS², Christopher Waples, PhD², and Janet Steele, PhD¹
Departments of ¹Biology and ²Psychology, University of Nebraska at Kearney, Kearney, NE, USA
Corresponding author: steelej@unk.edu

Abstract

Supplemental instruction (SI) has been shown to be effective in increasing student success in a wide variety of disciplines. Our study investigated the impact of the number of SI sessions attended on student success on exams and the effectiveness of remote SI compared to face-to-face (FTF) SI. Data were gathered for nearly 1,200 students enrolled in the first semester of a sophomore-level anatomy and physiology course at the University of Nebraska at Kearney (UNK). The number of SI sessions each student attended, if any, prior to each exam was compared to exam performance. Results for 2013 – 2017 demonstrated that attending even one SI session had a positive impact on exam performance, and an increase in exam performance was seen with additional SI attendance up to three sessions prior to each exam. We took advantage of the remote SI offered in the Fall of 2020, due to COVID-19, to investigate a potential effect of delivery format on SI effectiveness. There was no difference in exam performance for students attending SI FTF (2019) compared to students attending remote SI (2020), while attending SI in either format was associated with better exam scores. Our study is unique in examining the effectiveness of SI attendance at the level of individual exam performance and adds to the body of evidence that SI, whether FTF or remote, is effective in improving student success. https://doi.org/10.21692/haps.2022.013

Key words: anatomy & physiology, supplemental instruction, student success, online learning, exam outcomes

Introduction

Supplement instruction (SI), developed at the University of Missouri-Kansas City in 1973, is a peer-led collaborative learning support program built around course content. The program is designed as a proactive approach to historically difficult courses (International Center for Supplemental Instruction 2021). SI sessions are led by trained students who have previously taken the course. Because SI leaders have already demonstrated superior academic achievement in the course and attend lectures with the students currently enrolled, they are able to clarify lecture material and draw upon their previous experience in the course. SI leaders receive training and are supervised throughout the semester. They are trained in administering participative activities, questioning techniques, and quiz development that promote group work and peer learning while reducing test anxiety (Fayowski and MacMillan 2008). Strategies utilized in SI sessions are intended to be transferable to other courses and are beyond normal study skills (International Center for Supplemental Instruction 2021).

SI has been shown to be effective in a wide range of disciplines spanning undergraduate, graduate, and professional education (Dawson et al. 2014). For example, Fayowski and MacMillan (2008) found SI to increase the odds of success by 2.7 times for students in a first-year calculus course. Regular attendance at SI sessions during the first semester of an engineering program increased the chances of graduating by 20-35% (Malm et al. 2018).

For students enrolled in an introductory biology course, SI not only increased the percentage of students earning a grade of C- or higher from 73% to 85%, but the proportion of students who participated in SI and ultimately graduated was 67% compared to a graduation rate of only 59% for those students who did not participate in SI (Rath et al. 2007). In addition, Ning and Downing (2010) observed that SI had a significant impact on the academic achievement of first-year business students.

The impact of SI on underrepresented student groups is noteworthy. Bowman et al. (2021) analyzed results from 21 different courses across two semesters and noted that the strongest relationship between SI attendance, grade performance, and retention was for underrepresented student groups and for students who attended at least five SI sessions. Rath et al. (2007) observed that SI participation associated with an introductory biology course increased not only the percentage of students from underrepresented groups earning a grade of C- or higher from 51% to 76% but also the percentage of those students ultimately graduating from 50% to 73%. Bowman et al. (2017) and Malm et al. (2018) did not observe any gender-related differences in the effectiveness of SI, but Peterfreund et al. (2007) noted that the effect of SI was greater for males even though more females participated.

The COVID-19 pandemic forced colleges and universities to shift quickly to remote delivery of instruction. Faculty often
had to devise creative solutions for courses that had never before been offered remotely (Baldock et al. 2021; Coker 2020; Forster et al. 2020; Heiss and Oxley 2021; Van Heuvelen et al. 2020). Other student support services, including advising and tutoring, also had to move to remote delivery (Bouchez et al. 2021; Johns and Mills 2020). Online peer tutoring has been shown to be effective (Evans and Moore 2012; Gehret et al. 2017; Hrastinski et al. 2019). Hizer et al. (2017) examined the effectiveness of online SI delivery in four biology courses (Introduction to Molecular and Cellular Biology, Introduction to Experimental Design and Statistics, Molecular and Cellular Biology, and Genetics) and found that both online and face-to-face (FTF) students who participated in SI had higher course grades than those who did not attend SI at all. Meanwhile, Price et al. (2007) found online tutoring to be less effective than in-person tutoring, considering course grades, exam scores, and a questionnaire designed to determine tutor effectiveness. As such, it appears that online tutoring and SI can be effective; yet the previous research is somewhat conflicted.

Communication is an important aspect of proper SI implementation, and break-downs in communication between the parties involved, including students, SI leaders, course instructors, and SI coordinators, have been identified as potential impediments to successful SI implementation (Adebola 2020). Thus, online implementation of SI poses challenges to maintaining these lines of communication. In a post-COVID-19 world, where the use of online delivery formats in education has become commonplace, examination of the effectiveness of online SI is very relevant, and one concern of the present study.

Our project had four primary goals. We wanted to determine if: 1) students who attended SI sessions prior to an exam would score higher on that exam than those students who had not attended SI sessions, and 2) increasing the number of sessions attended was associated with additive increases in exam performance. We hypothesized that students who attended SI would have higher average exam scores (Hypothesis 1). Further, we hypothesized that the more sessions students attended prior to exams, the better they would do on the exams (Hypothesis 2). We also wanted to determine if: 3) attendance at SI sessions was associated with higher exam scores compared to non-attendance (whether FTF or remote), and 4) SI was more effective when conducted FTF than when conducted in online. We hypothesized again that those students who attended SI would score better on exams (Hypothesis 3). We also hypothesized that SI would be more effective FTF than online, with the FTF SI attendees scoring better than the students who attended SI online (Hypothesis 4).

**Methods**

SI was coordinated by the University of Nebraska at Kearney (UNK) Learning Commons—the tutoring center at our university—and was modelled after the SI that was first described and implemented at the University of Missouri-Kansas City (Martin and Arendale 1992). SI leaders were undergraduate students who had previously excelled in the course, were recommended by the course instructor, and were trained by the SI coordinator at the UNK Learning Commons. The SI leaders attended lectures, took notes, and planned and prepared activities for weekly SI sessions. These activities included games, topic discussions, and other collaborative activities, such as having the students in attendance role-play as instructors by taking turns re-teaching each other recent lecture material. Additionally, SI leaders met regularly with the course instructor and with the SI coordinator throughout the semester. SI leaders were compensated as student workers by the UNK Learning Commons.

Participation in SI was voluntary, and no extra credit or other incentive was offered. Students could attend SI as frequently or infrequently as they liked. As such, SI participation is self-selected and is an important factor to consider when drawing conclusions. This issue will be discussed at length below.

The project was approved by the UNK Institutional Review Board (UNK; IRB # 020321-1). UNK is a primarily residential undergraduate institution with a total enrollment (undergraduate and graduate) of approximately 6,400 students. Hispanic students comprise the largest minority population at approximately 15% of the undergraduate student body, and 86% of the student population is under age 25.

Participants in the study were enrolled in the first semester of a two-semester, sophomore-level anatomy and physiology course (BIOL 225) from fall semester 2013 through fall semester 2020. Over the study period 63.3% of the students enrolled in BIOL 225 were classified as sophomores (28-57 credits completed) with 22.8% classified as juniors, 9.3% as seniors, 3.6% as freshmen, and 1.0% as post-graduates. The course has a prerequisite of either completion of a college-level chemistry course or sophomore status. Students classified as “freshmen” were in their second year of university study but had not yet accumulated 28 credits. These students were permitted into the class to keep on track with their career goals. Fall semester 2018 was excluded from hypothesis testing, as SI was not available to students that semester. Enrollment in the course ranged from 194-224 students (average = 206 students) during the study period.

Our study focused on the impact of SI on individual exam performance and not the final overall grade in the class for two reasons. First, the composition of the individual exams continued on next page
was consistent from 2013-2020 while the composition of the final exam, which constituted approximately 20% of the final grade, changed significantly. Second, approximately 35% of the final grade in the class was from the laboratory component and SI did not address laboratory objectives.

Exams consisted of both multiple choice and short essay questions. For the comparison of in-person and remote SI, the exams had the same multiple-choice questions, which comprised 80% of each exam, and comparable essay questions. Students taking the exams in-person (2019) were proctored and had a time limit of 55 minutes. Students taking the exams remotely (2020) were required to use the Respondus Lock-Down Browser® with camera on and had the same 55-minute time limit. In each case the exams were closed book.

The data set was created by merging course records with SI attendance data collected and archived by the UNK Learning Commons. As such, the resultant data set included the date of each attended SI session for every student, along with exam scores. For students who took the course multiple times, only the first attempt was used for analysis. Similarly, data for students who took the course for the first time prior to the window of the study were not included in the study. Finally, students who withdrew from the course were not included.

In testing Hypotheses 1 and 2, the data set was derived from 860 students completing the course between 2013 and 2017. Across that span, 339 of the students chose not to attend the optional SI sessions. The remaining 521 (60.5%) students attended at least one session during the course and 82.0% of them attended up to as many as seven sessions throughout the semester.

Data used to test Hypotheses 3 and 4 were gathered for 326 students completing the course in 2019 and 2020. These two years were selected to allow for a comparison between traditional in-person (FTF) and online session formats, necessitated by the COVID-19 pandemic. In 2019, SI sessions were offered in-person, but in 2020, SI sessions were offered only in an online format (via Zoom). In 2019 (FTF) approximately 40% of students attended at least one SI session with 35% attending prior to exam 1, 22% between exams 1 and 2, 12% between exams 2 and 3, and 16% between exams 3 and 4. In contrast, in 2020 (remote) approximately 32% of students attended at least one SI session with 14% attending prior to exam 1, 19% between exams 1 and 2, 11% between exams 2 and 3, and 7% between exams 3 and 4.

All hypotheses were tested using linear mixed modeling to allow assessment of the relationships of interest. Linear mixed modeling is a hierarchically structured approach that allows for nesting individual participants into categories of SI session attendance, on a per exam basis, allowing for an aggregated assessment of the role SI attendance plays in exam performance. Two such analyses were conducted. The first tested Hypotheses 1 and 2, highlighting the impact of incremental SI attendance prior to each exam. The second tested Hypotheses 3 and 4, incorporating session format and its interaction with SI attendance prior to each exam. Rather than imposing theoretical structures, unstructured models were used to remove constraints on covariance matrices.

### Results

Hypotheses 1 and 2 were tested by examining the relationship between the number of SI sessions attended and subsequent exam performance using linear mixed modeling. The analysis revealed that number of SI sessions accounted for a significant proportion of variance in exam scores, $F(4,2660.13) = 16.59, p < .001, \eta^2_p = 0.01$ (Table 1). Post hoc analysis using a Bonferroni correction was conducted to identify differences in exam performance across the number of SI sessions attended (Figure 1). Pairwise comparisons indicated that students attending any number of SI sessions performed significantly better on the associated exam than did students who did not attend any such sessions ($p < .001$). Furthermore, there were instances of significant increases

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated B</th>
<th>S.E.</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>95% Conf. Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Intercept (Four or More Sessions Attended)</td>
<td>81.121</td>
<td>1.236</td>
<td>3200.210</td>
<td>65.611</td>
<td>---</td>
<td>78.697</td>
</tr>
<tr>
<td>One Session Attended</td>
<td>-5.430</td>
<td>1.223</td>
<td>2842.110</td>
<td>-4.439</td>
<td>&lt; .001</td>
<td>-7.829</td>
</tr>
<tr>
<td>Two Sessions Attended</td>
<td>-4.268</td>
<td>1.219</td>
<td>2700.087</td>
<td>-3.502</td>
<td>&lt; .001</td>
<td>-6.658</td>
</tr>
<tr>
<td>Three Sessions Attended</td>
<td>-2.066</td>
<td>1.332</td>
<td>2626.353</td>
<td>-1.551</td>
<td>0.121</td>
<td>-4.678</td>
</tr>
</tbody>
</table>

Note: Significance testing compares each number of sessions to “Four or More Sessions Attended,” which is represented by the intercept in this model.

**Table 1.** Estimates of the Effect of SI Session Attendance on Exam Performance.
in exam performance when students attended more sessions. Specifically, these differences were noted between attending a single versus three or four sessions, and between attending two versus four sessions ($p < .01$).

No significant difference in exam performance was found between those who had attended exactly three sessions and those who had attended four or more sessions ($p = 1.00$), indicating that optimal benefits for SI participation seem to emerge with three or more sessions attended. Hypotheses 1 and 2 were supported by the data.

To test Hypotheses 3 and 4, a second linear mixed model analysis examining the unique and combined relationships between SI attendance, course and session format, and exam performance was conducted. Individual participants were nested within both course format and attendance or non-attendance of SI prior to each exam. The analysis revealed that students attending SI prior to exams ($M = 78.97$, $SE = 0.74$) performed significantly better on those exams than did students who did not attend any such sessions ($M = 77.29$, $SE = 0.97$), $F(1,896.52) = 4.46$, $p < .001$, $\eta^2_p = 0.005$. Hypothesis 3 was supported by the data.

Hypothesis 4 was tested by examining the interaction between dichotomized SI attendance and session format on exam performance. The analysis revealed no significant interaction between the variables, $F(1,896.52) = 0.30$, $p = .587$, $\eta^2_p = 0.0003$. The non-significant interaction suggests that presentation format, online or FTF, did not moderate the effectiveness of SI sessions (Figure 2). Hypothesis 4 was not supported by the data.

Discussion

SI attendance was associated with higher exam scores over the course of the semester. This is consistent with previous research examining the effectiveness of SI on a variety of student outcomes, including course grades (Arendale 1997; Hizer et al. 2017), retention or continued enrollment (Bowman et al. 2021), probability of graduation (Grillo and Leist 2013; Paabo et al. 2003), and cumulative GPA (Grillo and Leist 2013; Ogden et al. 2003). Our data adds to the already strong body of evidence supporting SI as an effective program for student success.

One of the primary objectives of the current investigation was to determine how many sessions a student would need to attend to see the benefits of SI, as well as how much added benefit additional sessions would confer. This type of information is important from a logistical perspective, when SI coordinators and course instructors determine the number...
of SI sessions to offer. The way in which we approached the organization of our data set and the analyses utilized allowed us to address these questions directly. Our data set was structured so that we examined the effect that attending SI sessions prior to each exam had on exam performance. We found that the effect of SI on exam performance increased incrementally with additional pre-exam SI sessions attended. The effect of SI appeared to plateau at the three-session mark, after which no added benefit was seen. This would suggest that, if under logistical constraints from the instructor and SI coordinator end (e.g., finding enough SI leaders, funding for compensation), or those from the student end (e.g., making time to attend SI), three sessions prior to each exam might be sufficient. It is possible, however, that some students may still receive benefit beyond the three-session mark, a possibility which is discussed in more depth below. Furthermore, our data suggests that attending one session prior to each exam is sufficient to see an effect of SI on exam performance.

Our data resulting from testing Hypothesis 2 is consistent with previous research that has found the effect of SI to increase with increasing number of sessions attended. Hizer et al. (2017) determined that attending three sessions over the course of the semester was necessary for the effect of SI to manifest, while Fayowski and MacMillan (2008) reported five sessions to be necessary. Further, Bowman et al. (2021) and Kochenour et al. (1997) reported larger effects of SI as the number of sessions increased, although Bowman and colleagues noted a positive correlation when students attended only one session.

We also noted a significant effect of only one session, but it is important to keep in mind that we measured the effect at the exam level and not with an overall measure, such as course grade. Meanwhile, Grillo and Leist (2013) reported that the number of hours spent at academic support services, which included SI, were positively correlated with GPA and likelihood of graduation. Paabo et al. (2003) found that attending one SI session in a semester was not significantly associated with increased odds of graduation but attending more than two was. Finally, Arendale (1997) saw a positive correlation between the magnitude of the effect of SI and the number of sessions a student attended, with the trend levelling off around the 8-11 sessions-per-semester mark.

Taken together, the effect of SI becomes stronger with an increasing number of sessions attended, to a point, after which the trend plateaus. In fact, the trend is very consistent across studies. We found the plateau mark to be at three or more sessions per exam (i.e., 12 sessions over the course of the entire semester), which is similar to that of 8-11 sessions over the semester reported by Arendale (1997). However, the intricacies of this trend, such as the importance of regular attendance over the course of the semester, as opposed to cramming all the sessions in the last three weeks before the final exam, is less supported by evidence. Indeed, much of the previous research has examined the number of SI sessions at the semester-level. In the present study we examined the effectiveness of SI attendance at the level of individual exams. This gives our data set a level of temporal resolution such that we can confidently conclude that regular SI attendance prior to each exam is important for student success, as measured by individual exam performance.

In a similar study to our own, Price et al. (2012) examined whether peer-assisted study session (i.e., SI) attendance prior to individual quizzes in an introductory psychology course impacted quiz performance. Attendance was dichotomized, however, giving rise to just two groups—attendees and non-attendees, and the analyses were performed on just one semester’s worth of data. In general, those students who attended the study sessions scored better on the quizzes and the final exam than did students who did not attend the study sessions. The authors interpreted these results to suggest that peer-assisted study sessions are effective both in the short-term (i.e., individual quizzes) and the long term (cumulative final exam).

Meanwhile, Ogden et al. (2003) reported that SI participation in an introductory political science course was associated with gains in course grade, as well as overall GPA during the quarter of SI participation, but only for those students who were conditionally enrolled at the university. Conditional enrollment meant that these students were expected to meet a certain number of requirements before receiving full-enrollment status. Interestingly, the conditionally enrolled students who were attending SI were indistinguishable in terms of GPA from those traditionally enrolled students not attending SI. One way to interpret this result would be that SI has the capacity to help students who are lacking in college-preparedness to match the performance of their better-prepared peers. The authors also noted, however, that over time, and by the next year, the cumulative GPA of this group tended to drop, suggesting that perhaps an SI “booster” is necessary to see continued long-term benefits from the program. Taken together, it appears that SI has both short-term and long-term benefits, and as our data suggests, regular attendance appears to be very important.

Another key objective for the present investigation was to determine whether offering SI in an online format would be as effective as the traditional, in-person setting. First, we again predicted that both online and in-person SI would show significantly higher exam scores relative to no SI. This hypothesis was supported by our data, such that students who attended SI had significantly higher exam scores than those students who did not attend SI. After validating the SI treatment in these years, we could then examine the role of SI format on exam performance.

continued on next page
Communication has been identified as an important aspect of proper SI implementation (Adebola 2020). As such, we expected that channels of communication would be negatively impacted when SI was offered online during the fall of 2020 due to the COVID-19 pandemic. However, the format in which SI was offered, online or in-person, did not affect the effectiveness of SI in those semesters examined. This result was somewhat unexpected, given that tutoring was reported to be less effective in an online setting compared to a tradition in-person format (Price et al. 2007). Hizer et al. (2017), however, reported only minor differences in effectiveness between in person and online SI formats. This is consistent with the results that we report. Further, it is also possible that teachers, students, and SI leaders have become better equipped to teach, learn, and communicate in an online setting because of inadvertent practice in that setting stemming from the pandemic, which first moved education online in the spring of 2020.

In terms of limitations, we would be remiss to avoid discussion of an inadvertent problem with research involving SI. Namely, selection bias has been identified as a serious problem to studying the effects of SI. While differences can be seen between groups, it is hard to parse out selection bias to the SI group (Etter et al. 2000). Are the effects of SI attendance really all due to the intervention, or do other variables, such as motivation, explain the between-group differences? In other words, are students who choose to attend SI more likely to succeed, regardless of the SI intervention itself?

While previous research has shown that accounting for covariates (e.g., GPA, ACT scores, high school GPA) does reduce effect sizes, the effects of SI do persist (Bowman et al. 2021; Hizer et al. 2017). Furthermore, Hizer et al. (2017) concluded that there was no selection bias for SI attendance in their study, using GPA as a covariate, and Bowman et al. (2021) determined that ACT scores were similar for SI attendees and non-attendees, although SI attendees did tend to have higher high school GPAs. Grillo and Leist (2013), however, reported that high school GPA, SAT score, and ACT score were not significantly correlated with graduation likelihood in their study. Moreover, Fayowski and MacMillan (2008) still saw a medium effect of SI attendance on final grade in a calculus class after accounting for pre-GPA and gender.

Peterfreund et al. (2008) examined course grades for roughly 12,000 students in STEM courses over several years and found that SI participants did not have higher SAT scores or high school GPAs and tended to be worse in these areas. Further, Kohenour et al. (1997) found no correlation between high school GPA nor ACT scores and SI attendance, indicating that there is not a significant difference between students who attend SI and those who do not. These same researchers also found no correlations between number of hours that a student worked each week, nor the number of credit hours enrolled, and SI attendance. Similarly, Arenale (1997) found no differences between SI participants and non-participants over several demographic variables, including “gender, age, working status, high school rank, standardized test scores, ethnicity, or other factors” (p. 35). Finally, Malm et al. (2018) found the effects of SI attendance to persist after accounting for gender and prior academic achievement. Taken together, the data cited here suggest that the effect of SI attendance, while perhaps moderated by other factors, is indeed an effect of the SI intervention.

Meanwhile, Guarcello et al. (2017) used coarsened exact matching to minimize selection bias. The method matches students in the SI group to students in the non-SI group based on a series of co-variates. SI participation was then assessed as a function of likelihood of passing the course. The SI group was 2.2 times more likely to pass the course with a C or better. The authors did report that, prior to addition of the coarsened exact matching model, the SI group scored better for three of the four exams in the course. After the model was added, significance was only reached for one of the exams. While the effect size was reduced, this analysis further bolsters the conclusion that much of the positive student outcomes associated with SI attendance can be attributed to SI itself.

In the present study we did not account for any factors other than the SI intervention, and thus our results must be interpreted with caution. Yet, as has been discussed above, previous research would suggest that much of the effect of SI seen is more than likely attributable to the SI intervention. Regarding motivation specifically, Arenale (1997) reported that motivation plays a role as highly motivated students performed better than students with less motivation. Still the bulk of the effects seen with SI attendance were not explained by motivation.

Furthermore, it appears that students high in self-efficacy and students low in self-efficacy attended SI at higher rates than those students somewhere in the middle (Price et al. 2012). This complex relationship between self-efficacy and SI attendance could perhaps explain the plateau in SI benefits after a certain number of sessions that we and other researchers have reported. Indeed, Arenale (1997) discussed that, upon talking with those students who attended the maximum number of sessions, it was determined that that group contained a subset of students who had intentions of withdrawing from the course but were seemingly able to push through to the end by attending many SI sessions. It could be that the students who attend very high numbers of SI sessions are indeed benefiting, and perhaps for some of these students, high levels of SI attendance could be protecting against course withdrawal. As we did not include any such measures nor had any way to talk with students about their individual experiences, our interpretation is limited to speculation.
Despite the limitations discussed above, there are some important strengths to our study. Our analysis included several years of student data with consistent patterns across all four exams, indicating that the effect of SI at our institution is robust and temporally stable. Along with this, the anatomy and physiology course was taught by the same instructor for the duration of the study period, with the course format, exam format, and course content remaining very consistent. This gives continuity to our dataset and made it easy and logical to coalesce the data from across years into one master data set for analysis. Further, the detailed records kept by the UNK Learning Commons, when merged with course exam data, allowed for a very detailed examination of SI attendance and subsequent exam performance. Thus, our data set boasts a high temporal resolution regarding SI attendance. For this reason, we can draw conclusions regarding the importance of regular attendance of SI on individual exam performance, over the course of the semester.

Meanwhile, the COVID-19 pandemic and subsequent shift to a remote learning space allowed for a serendipitous opportunity to examine the effectiveness of online SI delivery in Hypothesis 4. Given the results of Hypotheses 1, 2, and 3 we were able to, with confidence in the effectiveness of SI at our institution, compare the traditional in-person SI format in 2019 to the online format in 2020. This allowed us to directly address the question of effectiveness of online SI, which is of especially high relevance in a post-COVID-19 collegiate education system.

Conclusion

SI has been utilized to aid in undergraduate education for nearly five decades. Much research over that span of time has found SI to be an effective program to aid in student success, and the data we present here strengthens that conclusion. Our investigation also offers insight as to the relationship between the number of sessions attended and the benefit conferred from them, as well as regarding the effectiveness of SI in an online format. Specifically, the results of Hypotheses 1 indicate that students who attended SI prior to any given exam scored significantly better on that exam than students who did not attend SI. Further, in testing Hypothesis 2 we determined that the effect of SI increased in magnitude with the number of sessions attended prior to each exam, up to three or more sessions, and this trend was consistent and significant across all 4 exams. Finally, we were also able to test the effectiveness of SI in an online format. The results of Hypotheses 3 and 4 found that the effect of SI on exam performance was not significantly moderated by format, whether in-person or online. As such, at our institution, online SI delivery was as effective as in-person, FTF SI delivery.

Acknowledgments

The authors would like to thank Patrick Hargon and Emily Bahr of the UNK Learning Commons for providing data necessary for this study.

About the Authors

Blase Rokusek completed his BS in Psychobiology from UNK in 2021 and is currently working on his master’s degree in Biology at UNK. Emilee Moore also graduated from UNK with a BS in Psychobiology in 2021 and is currently working on her doctorate in Occupational Therapy at St. Ambrose University. Christopher Waples, PhD, is an associate professor of Psychology at UNK. He teaches courses in statistics and research methods and his research interests include measurement and performance feedback. Janet Steele, PhD, is a professor of Biology at UNK. She teaches undergraduate anatomy and physiology and directs online MS programs in STEM Education and Health Sciences.

Literature Cited


continued on next page


continued on next page


---

**Become a Member of HAPS Today!**

The Human Anatomy & Physiology Society (HAPS) is dedicated to promoting excellence in the teaching of Anatomy and Physiology at college, universities and related institutions.

- Connect with colleagues also pursuing success in teaching A&P
- Discounted rates for annual and regional conferences
- Access Teaching Tips for tough topics
- HAPS Institute short courses for ongoing professional development
- Open access to our peer reviewed journal, *The HAPS Educator*
- Grants and scholarships for research and travel
- Additional member resources listed [here](#)

For more information, contact HAPS at [info@hapsconnect.org](mailto:info@hapsconnect.org) or at 1-800-448-4277. Follow [this link](#) to join and start benefiting from membership today!
What the Flip? A Pilot Study Perspective on the Flipped Classroom for Medical School Physiology

Chasity B. O’Malley, PhD
Kiran C. Patel College of Allopathic Medicine, Health Professions Division, Terry Building, 3200 South University Drive, Nova Southeastern University, Fort Lauderdale, FL 33328
comalle0@nova.edu

Abstract

Flipped classrooms have been shown to improve engagement in undergraduate classrooms, in general, but there have been fewer studies reporting on how they are received in the undergraduate medical education classroom. This perspective piece reflects on one instructor’s experience with flipping the classroom as a pilot study when teaching medical students. Some preliminary survey data reveals student feedback with regard to the effectiveness of the flipped classroom sessions for first year medical students. While the response rate was very low for the survey, it still provided an indication of the usefulness of the sessions, as well as some helpful open-ended feedback from the participants. This pilot study will be repeated in the Spring 2023 semester. https://doi.org/10.21692/haps.2022.006

Key words: flipped classroom, medical education, physiology, engagement

Introduction

Traditional didactic lectures are a standard of content delivery for many medical schools. However, in recent years, there has been a push to include more active learning in undergraduate medical education. The active learning techniques typically include case-based learning, team-based learning, problem-based learning, flipped classroom, and small group projects/discussions (Ramnanan and Pound 2017). Other active learning techniques can include think-pair-share, one-minute papers, and check out quizzes (Lord et al. 2012).

The flipped classroom is a pedagogical technique that involves using class time for the processing parts of learning rather than the initial learning parts (Brame 2013). The flipped classroom can take many forms, but it often involves providing students with prescribed educational content in the form of a reading, a video, or some other form of preparation. The students preview this material prior to coming to class, allowing them to arrive with a level of understanding of the content to be covered in the class session which now consists of activities dedicated to enhancing application of the students’ knowledge through discussion and collaboration with each other, with guidance from the instructor as needed (Ramnanan and Pound 2017).

There has been an increase in the use of the flipped classroom in medical education, with positive correlations made to the effectiveness of the approach for promoting knowledge acquisition in comparison to traditional didactic lectures (Chen et al. 2017). Additionally, the flipped classroom has been shown to increase engagement and satisfaction with content delivery in third year medical students (Chowdhury et al. 2019). This perspective focuses on a pilot study using the flipped classroom in a systems-organized block course in undergraduate medical education.

Methods

Description of the Intervention

The goal of the flipped classroom pilot study reported here was to increase student engagement and satisfaction during the lecture sessions. The student population consisted of students enrolled in the Gastrointestinal, Human Nutrition, Endocrine, and Reproduction (GIHNER) block course. There were 51 students in total divided equally between males (49%) and females (51%). The average age was 24 years and 7.8% of the cohort had completed a master's degree prior to enrolling in medical school. From the admission record, 13.7% were Hispanic/Latinx, 11.8% African American, Black and/or Afro-Caribbean, 31.4% Asian, 41.2% White, 11.8% “other”, while 11.8% did not reveal their ethnicity upon admission. GIHNER is a ten-week, block-style course, categorized as the third course in basic science education for first year medical students. It is preceded by the Professional Immersion course, the Fundamentals (overview) course, and Hematology. The GIHNER course consists of non-mandatory lectures and mandatory problem-based learning (PBL) sessions that focus on anatomy, biochemistry, clinical applications, immunology, microbiology, nutrition, pathology, and physiology. There are also mandatory laboratory sessions for anatomy and histology that promote hands-on learning.

For the non-mandatory lecture components, teaching faculty were encouraged to use active learning techniques. Most sessions were fifty minutes in length and were held in a lecture hall. For most faculty, active learning included
a sprinkling of questions throughout the session where the students could verbalize a response or use an electronic polling system to answer. For prework, students were given a variety of resources including pre-recorded lectures, videos explaining the topics to be covered, readings in the textbook, and/or journal articles related to the topics.

For the active learning sessions of this course that I taught, a mix of lectures with questions throughout and the flipped classroom approach was used, resulting in seventeen total fifty-minute sessions. To increase student engagement, the classroom was flipped for seven of the seventeen sessions for this pilot study. The seven sessions were the first seven that I delivered to the students and contained all but two of their endocrine physiology topics.

For the flipped classroom, a voice over recording was done on a PowerPoint presentation of a traditional didactic lecture and then posted for the students as a pre-session assignment, along with suggested textbook reading. During the fifty-minute in-class session, questions related to the material were discussed. The questions were a mix of those created by the author for the session and those taken from board examination preparation materials available through the university library. Using Bloom’s taxonomy, the question difficulty ranged from low level (recall) to high levels such as asking the student to create a suitable treatment plan for the “patient” (Anderson et al. 2001).

The questions were presented in Poll Everywhere (www.polleverywhere.com/) to encourage students to answer anonymously, and then the class was given the opportunity to compare and discuss the answer choices. These in-class deliberations with peers led to some robust discussions about why certain answer choices were correct or not correct and guided the students to arrive at the correct answers with very little intervention from the instructor.

Survey Method

Student perceptions of the flipped classroom sessions were collected with a thirty-question anonymous survey developed by Aljaraideh (2019), which was supplemented with several additional questions to tease out unique aspects related to our course. The additional questions are shown in Table 1. A Likert scale of 1 (strongly disagree) to 5 (strongly agree) was used in this survey, which was created in Microsoft Forms and distributed as a link both in class and through an email sent to all students in the course, since everyone had the opportunity to view the recorded flipped classroom sessions, whether they had attended in person or not.

This project was approved as exempt (project # 2022-86-NSU) by the Institutional Review Board of Nova Southeastern University, and informed consent was obtained from all participants who completed the survey.

Results

Unfortunately, only 2 participants filled out the survey for the pilot study, despite the fact that all 51 students received an initial invitation to complete the survey as well as a reminder. The preliminary results of the survey are mostly positive with both students liking the flipped classroom approach and indicating that it enhanced their learning (Table 2).

<table>
<thead>
<tr>
<th>Multiple Choice Questions</th>
<th>Free-response Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>For Dr. O’Malley’s lectures, do you prefer flipped classroom or standard didactic lectures?</td>
<td>Which topics would you prefer to see as standard lectures?</td>
</tr>
<tr>
<td>A. Flipped</td>
<td></td>
</tr>
<tr>
<td>B. Standard didactic</td>
<td>Do you have any comments on what went well in the flipped classrooms?</td>
</tr>
<tr>
<td>C. Depends on the topic</td>
<td>Do you have any suggestions for improvement for future sessions?</td>
</tr>
</tbody>
</table>

Table 1. Additional Questions for the Flipped Classroom Survey
<table>
<thead>
<tr>
<th>Question</th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel that watching videos and taking notes contribute efficiently to my learning.</td>
<td>4.0 ± 1.41</td>
</tr>
<tr>
<td>With flipped classroom model, I feel more prepared for my exam.</td>
<td>3.0 ± 1.41</td>
</tr>
<tr>
<td>I like watching the lessons on video.</td>
<td>4.0 ± 0</td>
</tr>
<tr>
<td>I try to learn as much as possible while watching the videos.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>I wish more instructors use the flipped or inverted classroom model.</td>
<td>5.0 ± 0</td>
</tr>
<tr>
<td>I frequently pause or repeat parts of the videos in order to increase my understanding of the material.</td>
<td>5.0 ± 0</td>
</tr>
<tr>
<td>Flipped classroom encourages me to practice critical and creative thinking.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>Learning foundational content prior to class greatly enhances my understanding of material.</td>
<td>3.5 ± 2.12</td>
</tr>
<tr>
<td>Flipped classroom gives me the opportunity to ask more questions inside the classroom.</td>
<td>3.5 ± 0.71</td>
</tr>
<tr>
<td>Flipped classroom attracts my attention to learning and teaching process.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>With flipped classroom, we have to do more work out of the classroom.</td>
<td>5.0 ± 0</td>
</tr>
<tr>
<td>Flipped classroom can be a suitable teaching strategy.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>Flipped classroom can improve interest in exploring topics.</td>
<td>3.5 ± 2.12</td>
</tr>
<tr>
<td>I felt prepared to complete course tasks in class after listening to the video content.</td>
<td>3.0 ± 1.14</td>
</tr>
<tr>
<td>Flipped classroom is more engaging than the traditional classroom.</td>
<td>5.0 +/- 0</td>
</tr>
<tr>
<td>Flipped classroom gives me less class time to practice the concepts of course.</td>
<td>3.0 ± 1.41</td>
</tr>
<tr>
<td>Flipped classroom reduces the effort to understand the basic knowledge of the subject matter.</td>
<td>2.0 ± 1.41</td>
</tr>
<tr>
<td>Flipped classroom, along with delivery of content outside class and problem solving in class, is an instructional method appropriate for my specialization.</td>
<td>5.0 ± 0</td>
</tr>
<tr>
<td>I am more motivated to learn the concepts of course via the flipped classroom.</td>
<td>4.0 ± 1.41</td>
</tr>
<tr>
<td>Flipped classroom improved collaborative learning.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>Flipped classroom can improve interest in class.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>I got the ability to self-pace my learning with flipped courses.</td>
<td>3.5 ± 2.12</td>
</tr>
<tr>
<td>Flipped classroom gives me greater opportunities to communicate with other students.</td>
<td>2.5 ± 0.71</td>
</tr>
<tr>
<td>I believe that I am able to learn material with flipped classroom instruction better than with traditional lecture-based instruction.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>I would recommend flipped classroom to a friend.</td>
<td>5.0 ± 0</td>
</tr>
<tr>
<td>Flipped classroom matches my learning style.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>I feel that mastering learning through flipped classroom improved my academic achievement.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>Flipped courses did not limit my interaction with instructors.</td>
<td>5.0 ± 0</td>
</tr>
<tr>
<td>I feel that mastering learning through flipped classroom improved my course understanding.</td>
<td>4.5 ± 0.71</td>
</tr>
<tr>
<td>Flipped classroom learning has reduced my dependency on the instructor.</td>
<td>3.5 ± 2.12</td>
</tr>
</tbody>
</table>

*Table 2. Results from the Flipped Classroom Survey (2 respondents)*
For the free-response questions, both students stated that they preferred the flipped classroom to the standard didactic lecture. There were no specific topics mentioned that would be better flipped than standard didactic; although it was stated that “most” topics would be better as flipped. Additionally, one comment indicated that their engagement was higher, likely due to the pre-recorded lecture as preparation for the session: “I think the standard lecture prior to flipped classroom sessions increases level of engagement.” It was also stated by a student that the flipped classroom helped “… students to communicate what they learned and build upon that knowledge……”. This statement supports the idea that the flipped classroom focuses on the application side of learning and helps the students to take the additional step to understand the topics they are learning.

**Discussion**

Overall, the flipped classroom sessions were well received by the students. Comments from students in previous cohorts on my teaching performance reviews stated clearly that they liked the flipped-classroom question sessions. As far as verbal feedback from the students, they indicated that they loved the opportunity to think about the material in a friendly environment where their peers or instructor could correct misconceptions and encourage them to go to the next level with their thought processes. They think more about “what if this happens” rather than “what is happening”. Additionally, I noticed that there was an increase in student attendance during the current year at the non-mandatory lectures that consisted of a flipped classroom. In a traditional lecture, we may get 7 students in attendance, while the flipped classrooms were often associated with attendance by 15-20 students. For the future study, attendance at the sessions will be tracked to determine the extent of this observation.

As an area for improvement, on the survey question related to how the flipped classroom could be improved, the statement was made that there “…should be a logical flow to the flipped classrooms that builds our knowledge in a progressive way. Through lectures, depth can be explored but with flipped classroom ….. it seems we are thrown into the deep end every time and I would like to see us ease into the deeper topics with a flipped classroom.” This comment likely came from sessions presented by a colleague of mine whose approach to the flipped classroom was heavily centered on clinical applications and used the Socratic approach to questioning. My sessions had a flow from very basic to more applied and thought-provoking questions as the session progressed. Moving forward, the plan is to work with the other faculty to refine the approach to the flipped classroom so that a consistent pattern is followed.

Many faculty are hesitant to approach the flipped classroom as a method of material delivery. When discussing with colleagues, reasons for this hesitation included issues such as not knowing enough about how to do it, not being confident enough that the students would prepare in advance, and/or not having the time to invest in preparing for the session in this way. I encourage all faculty to try this method out for just one session. Once they become comfortable with it and the students know the expectations for an appropriate level of preparedness, the flipped classroom is a wonderful method for content delivery. As shown by the limited survey results, the flipped classroom encourages engagement more than a traditional didactic lecture and also is something that the students wish more faculty would do.

The pilot study will be repeated in the Spring 2023 semester with more advertisement for the survey and a longer window for the students to participate. The future plan is to include a link through a QR code to the survey at the end of each of my sessions rather than waiting until the end of the block for feedback.

**About the Author**

Chasity O’Malley is an Associate Professor of Medical Education and Physiology at Nova Southeastern University. Her research goals aim to improve the learning experience for students by helping them learn to study and interact with the material and for faculty by helping guide them on implementing active learning into their classrooms. She is a co-principle investigator for an NSF grant related to this work. She also is actively involved in promoting diversity through her funded research projects centered around enhancing training for medical students related to the LGBTQ population.

**Literature Cited**


continued on next page


Ten Years in the Human Anatomy and Physiology I Classroom: A Retrospective Analysis of Student Preparation, Engagement, Performance, and the Impact of COVID-19

Carol A. Britson, PhD
Department of Biology; University of Mississippi; University, MS 38677; cbritson@olemiss.edu

Abstract
Reflections on the efficacy of pedagogical changes and practices and their effect on student performance are often hindered by incomplete data, small sample sizes, and the confounding variables of multiple instructors and teaching sites. Observations from such retrospective analyses, however, are highly sought after by instructors and administrators interested in what methods significantly enhance student learning and comparisons of student success across instructors and institutions. Compilation of student data from ten years of Human Anatomy and Physiology I at the University of Mississippi enabled statistical analyses of how changes in course design over ten years of instruction, including remote instruction during the COVID-19 pandemic, were associated with student engagement and performance in Human A&P I with a large data set (n=3305) from students taught by a single instructor. Univariate analysis of variance, bivariate correlation, and discriminant function analysis (DFA) tests revealed multiple significant differences over time. Specifically, the DFA indicated that 89.5% (Discriminant Function 1) of the variation in overall course performance (i.e., letter grade) is explained by student performance variables of exam average, lab practical average, lab quiz average, and the number of Supplemental Instruction (SI) sessions attended. For Discriminant Function 2, 8.1% of the variation is explained by student engagement variables of the number of missed lecture assignments, lab assignments, and online assignments. Institutionally, these results will be used to continue effective course practices, identify engagement strategies that enhance student motivation and reduce anxiety, and develop a performance dashboard that will both identify struggling students and coach students towards success in A&P.

https://doi.org/10.21692/haps.2022.010

Key words: anatomy, physiology, human, education, active learning, laboratory, prerequisites, performance, engagement, COVID-19, pandemic, meta-analysis, retrieval practice, backwards design, course blueprinting, high-structure

Introduction
Milestones in an academic career can be as celebratory, and provoking of self-reflection and assessment (Seithers et al., 2021), as personal milestones. The completion of the Fall 2020 academic semester also marked the completion of ten years of my teaching Human Anatomy & Physiology (A&P) I at the University of Mississippi. These ten years saw a 25% increase in course enrollment combined with a campus-wide increase in careers in the health professions (U. Miss, 2014), implementation of a Supplemental Instruction program (U. Miss, 2022; Eroy-Revels et al., 2019), an increased role of undergraduate teaching assistants in the laboratory (Hopp et al., 2019; Luckie et al., 2020), use of evidence-based teaching practices including emphases on time management and self-disciplined study, and ended with a global pandemic disrupting nearly all components of higher education (Schaefer, 2022).

In 2011 my charge was to assume control, as the sole instructor, of the Human A&P program and teach the two, sequential courses. These two courses, Human A&P I and II are challenging and have high attrition rates; Lunsford and Diviney (2020) describe Human A&P as a challenging, entry-level course that is often viewed as a “killer” course by students. Students often enroll in A&P under-prepared, relying on study habits that served them well in high school incorrectly assuming that college is like high school and the responsibility for their learning is the instructor’s rather than theirs (Lunsford and Diviney, 2020). These factors lead to a high attrition rate and a risk of students having their goals dashed and leaving higher education with large debts and diminished prospects (Lunsford and Diviney, 2020). With a goal of improving positive student outcomes (e.g., engagement and enjoyment as well as increased performance) while reducing negative outcomes (e.g., frustration and complacency as well as decreased performance), the University of Mississippi courses have changed to reflect needs of both the course curriculum and student goals, limitations of course administration (e.g., academic calendar, lecture and laboratory space, etc.), and limitation of course (e.g., laboratory equipment) and student resources both financial and temporal. Using the high-impact, pedagogical techniques of course blueprinting (Coderre et al., 2009; Villarroel et al., 2018), backwards design (Wiggins and McTighe, 2005; Emory, 2014), and
high-structure (Wilton et al., 2019; Beck and Roosa, 2020) in conjunction with careful selection and development of learning activities and assessments and regular opportunities for retrieval practice (Bae et al., 2019; Dobson et al., 2017; Ritchie et al., 2019), has led to subjective year-to-year improvements in outcomes but analyses had never been undertaken to identify long-term, statistically significant changes.

Examination of attrition, or DFW rates (i.e., percent of students not meeting the requirement of passing A&P I in order to enroll in A&P II), and identification of at-risk students is common because of its important for students, educators, and administrators alike and has been the focus of several reports. Russell et al. (2016) found that factors associated with a higher pass rate included higher age (of the student); not taking a developmental reading, writing, or math class; taking any college credits prior to A & P I; not repeating any college course with or before A & P I; not being a first-generation student; taking a freshman biology, chemistry, or physics class before A & P I; and taking a daytime class. Sturges et al. (2016) found positive correlations between motivation, hours of study, and undergraduate grade point average and passing rates. Keller and Hughes (2021) examined the relationship between declared major and success in A&P I and found no significant differences in a two-year study. Young et al. (2019) found a significant, positive relationship between self-efficacy and gender (female). In an analysis of prior coursework and success in A&P I, Hopp (2009) found that a positive relationship between previous coursework in chemistry and success in A&P I.

These reports have both strengths and weaknesses. Russell et al.’s (2016) large sample size (over 3500 students) and length of study (13 years) is countered by variability in course resources (e.g., texts, lab manuals), faculty, campus, and grading cutoffs. Similarly, Sturges et al. (2016) has a large sample size (over 1200 students) but draws its data from both A&P I and II courses taught by multiple instructors over a two-year period. Young et al.’s (2019) data are based on a single semester’s study. Small sample sizes and limited time frames of study make inferences on factors associated with passing rates in A&P I challenging to produce. In 2016 the Human Anatomy and Physiology Society undertook an investigation into high attrition rates and the impact of prior coursework (e.g., prerequisites) on success in Anatomy and Physiology courses (Hull et al., 2016). The survey was broad in the scope of questions while narrow in the response variable (attrition) but failed to obtain enough responses (an estimated 220) to be able to perform statistical analysis to detect treatment effects while minimizing the potential for Type I and Type II errors (Jackson, 2017). While the present study only focuses on one Human Anatomy and Physiology course and cannot take the place of broad sampling of instructors and institutions, the sample size is large, and consistency of instructor and course grading scales enables greater insight into success rates in A&P I.

Institutional policy (e.g., contact hours, delivery format) and structural resources (e.g., lecture and lab room sizes and availability) may constrain development of the ideal A&P course experience, but incorporation of Human Anatomy and Physiology Society (HAPS) Learning Outcomes (HAPS 2019a) with backwards design and course blueprinting principles can create an enriching and rewarding experience for students and educators. The backwards design framework starts with identification of learning outcomes and supports curriculum planning and alignment to achieve congruence between student outcomes and measurable learning verbs or “facets of understanding” [(e.g., explanation, interpretation, application, perspective, empathy, and self-knowledge)] Wiggins and McTighe, 2005; Emory, 2014]. Identification of assessments appropriate to the desired learning outcomes is the second step of backwards design and can be simultaneously incorporated with course blueprinting principles. While the primary function of a course blueprint is to validate assessment tools, it can also be used to guide the selection of learning activities (Coderre et al., 2009; Villarroel et al., 2018). A course blueprint can be constructed methodically and quantitatively enabling objective insight to the relative weighting of course assessments (McDonald et al., 2016) and should be maintained through a systematic monitoring of course content (Coderre et al., 2009; Villarroel et al., 2018).

With practical applications for supporting the success of underrepresented minorities in introductory biology courses, a high-structure course design (Wilton et al., 2019) can also be used for human A&P courses. Courses with increased structure incorporate multiple formative assessments or interventions, including pre-class videos/reading coupled with quizzes, in-class active learning modules, mock exams, and small near-peer collaborative workshops to effectively promote student performance (Wilton et al., 2019). Bae et al. (2018) found that free recall and practice quizzing were the most singularly effective forms of retrieval practice (as compared to test generation and keywords) but combining practice quizzing and test generation with free recall led to significant gains in performance.

In practice, a high-structure design can model the self-discipline and time management skills that first and second-year undergraduate student lack (Lunsford and Divinney, 2020), and the visual and conceptual content within an A&P course provides multiple opportunities for retrieval practice and formative assessments. Implementation can be free-form and one-on-one, or small group, methodical (daily, weekly, etc.) and quantifiable using internet or publisher-provided resources. The latter provides convenience to the instructor, particularly with large classes, of synchronously entering scores into a learning management system.

continued on next page
However, changes in the curriculum of a course and its administration are rarely implemented solely for convenience, but rather are done to increase student proficiency in engagement and learning outcomes achieved. Empirically demonstrating that these objectives have been met is challenging due to small sample sizes; multiple instructors with changing instructional responsibilities; increased demands on both students and instructors; interaction of factors affecting student achievement; factors outside the influence of students or instructors; and inconsistent record keeping. This study is uniquely positioned to minimize these challenges of demonstrating positive, long-term impacts of curricular changes due to the following features: a single individual (Britson) was the instructor of record; lectures met three times per week at the same week days and times; laboratory sections met once per week for 2 hours on the same days and times; the sequence of lecture and laboratory topics remained the same; and the number of major assessments (lecture exams and laboratory practicals) remained nearly unchanged during the study period (2010-2020). The exceptions were due to the COVID-19 pandemic and the home institution shortening the fall 2020 semester by 1.5 weeks resulting in an altered lecture sequence and number of lecture exams. Diligent record keeping and the construction of a cumulative data set during these ten years presented an opportunity to combine reflection with statistical analyses, and comparison of a pandemic altered curriculum to the unaltered curriculum, both powered by data from over 3300 students. With a large data set, both Type I and Type II statistical errors can be avoided with increased sensitivity (Kaplan et al., 2014; Faber and Fonseca, 2014). Calculation of effect sizes (Cohen, 1988; Lakens, 2013) will also enable a reasoned interpretation of results [(i.e., determination of differences between statistical significance and real-world, classroom significance (Britson, 2020)).

The primary objective of this retrospective analysis was to examine how changes in course design over ten years of instruction were associated with student engagement and performance in Human A&P I. Longitudinal trends will be used to evaluate the impact of difference course delivery on students during a full semester of remote instruction due to the COVID-19 pandemic. In addition to contributing to the development of an A&P success dashboard that would potentially be useful to both instructors and students, the data set, in its size and breadth, can be a useful tool for other instructors wanting to compare their courses and data.

**Materials and Methods**

Human Anatomy and Physiology (A&P) I at the University of Mississippi is a 4-credit, combined lecture-lab course and is continued by Human A&P II which is also 4-credit, combined lecture-lab course. Human A&P I centers around the Human Anatomy and Physiology Society Learning Outcome Modules A through H (Body Plans through Nervous System; HAPS 2019a) and Human A&P II continues with Modules I through R (General and Special Senses through the Reproductive System; HAPS 2019a). Both courses are designed to meet the needs of students in the general education curriculum and in the allied health fields. Human A&P I is offered during the fall and first summer sessions while Human A&P II is offered during the spring and second summer sessions. Separate, advanced human anatomy, comparative anatomy, histology, and embryology courses are offered for students majoring in the sciences.

Students enrolled in Human A&P I at the University of Mississippi during the fall semesters of 2011-2020 were the subjects (n=3305) from which data were collected for this study [approved as Exempt under 45 CFR 46.101(b) (#4); University of Mississippi Institutional Review Board Protocol #21x-314]. Students in their 2nd year of study represented 44.4% of the group; students in their 3rd year, 22.7%; students in their 4th year, 17.4%; students in their 1st year, 12.1%; and students in their post-undergraduate education, 3.3%. Over 75 declared programs (i.e., majors) were represented with the largest groups coming from students majoring in exercise science (27.8%), allied health (pre-nursing, occupational therapy, cytotechnology, dental hygiene, health information management, radiological science, medical technology, etc.; 26.3%), general studies (10.6%), dietetics and nutrition (8.3%), biological sciences (7.4%), pre-pharmaceutical sciences (4.8%), and chemistry (2.1%). Common career goals included nursing, physician assistant, physical therapy, medicine, and dentistry (Hillhouse and Britson, 2018; O’Connor and Britson, 2017). De-identified, prior preparation data [highest ACT score, high school grade point average (GPA), and undergraduate GPA] for students enrolling for the first time in Human A&P I for the study period were obtained from the universities office of Institutional Research Effectiveness and Planning (IREP). Within the prior preparation data, sample sizes were unequal across years because the students may not have submitted a score, may have entered as a transfer, or may have just started their undergraduate education. The IREP office refrained from linking the data to individual students without written consent from each student. As this report is a retrospective study, these consent forms were not available.

Throughout the ten-year time frame of this study, all Human A&P I students were taught by the same instructor (Britson) in a single lecture section, and in the same auditorium, meeting at 8am on Mondays, Wednesdays, and Fridays during the fall semesters. Best practices for A&P instruction in large classrooms (e.g., student assistants, personal response systems, active learning, attendance tracking when feasible; Hill et al., 2017; Dogrell, 2021; Marwaha et al., 2021) were used throughout the study. Students met in groups of 30 for 2-hour lab sessions once per week.
Curricular modifications throughout the study are mapped by year in Table 1. The earliest changes included the adaption of a virtual cadaver resource for self-study and lab preparation (Kerce, 2013), and the reduction in the Teaching Assistant (TA) to student ratio in the lab from 1:30 to 1:15. This reduction enabled lab instruction to be more aligned with HAPS course and safety guidelines (HAPS, 2019b) and used an ‘experienced TA paired with a new TA’ model (Hopp et al., 2019) to enhance instruction and student engagement as well as curriculum continuity from year to year. Three to six, peer-led Supplemental Instruction (SI) sessions have been offered each semester throughout the study period. Weekly attendance was recommended throughout the study but was additionally encouraged by offering a small amount of extra credit (typically no more than 1 percentage point) starting in 2014. This incentive was part of an “extra credit bucket” of a variety of activities [e.g., volunteering for student research projects (Hillhouse and Britson, 2018; O’Connor and Britson, 2017), participating in course surveys, completing additional online homework, etc.] students could pursue to add up to 3-4 percentage points to their score at the end of the semester.

<table>
<thead>
<tr>
<th></th>
<th>Fall 2011</th>
<th>Fall 2012</th>
<th>Fall 2013</th>
<th>Fall 2014</th>
<th>Fall 2015</th>
<th>Fall 2016</th>
<th>Fall 2017</th>
<th>Fall 2018</th>
<th>Fall 2019</th>
<th>Fall 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Cadaver</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra Credit Bucket</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offered (Including for SI</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>attendance)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory TA to student</td>
<td>1:30</td>
<td>1:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade Equity</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization on Lab</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practicals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tortora and Derrickson 2011,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory Manual</td>
<td>Allen and Harper 2011,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-Risk Bonus Questions</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Included in Final Exam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item Analysis after each</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lecture exam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graded Online Homework</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (Mastering A&amp;P, Pearson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education, 2021a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Study Plans Provided</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Banks Provided</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with Lab Practicals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Questions Per</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture Exam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tracked lecture attendance</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(In Person)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture Quizzes</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Lecture Exams</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Students Completing the Course</td>
<td>283</td>
<td>353</td>
<td>345</td>
<td>362</td>
<td>348</td>
<td>355</td>
<td>349</td>
<td>327</td>
<td>342</td>
<td>243</td>
</tr>
</tbody>
</table>

Table 1. Implemented course design modifications in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020.
Ten Years in the Human Anatomy and Physiology I Classroom: A Retrospective Analysis of Student Preparation, Engagement, Performance, and the Impact of COVID-19

Several course changes, including a change in lecture textbook and lab manual, took place in the fall 2016 semester marking a shift to a “high-structure” course design (Wilton et al., 2019; Beck and Roosa, 2020) to enhance student engagement and performance. Multiple (i.e., 2-4 per week), low-stakes online homework assignments were added to the course to facilitate retrieval practice, an effective strategy to promote anatomy recall (Dobson et al., 2017; Bae et al., 2019; Ritchie et al., 2019). Assessments were developed and selected according to backwards course design principles (Wiggins and McTighe, 2005), and weekly study plans (e.g., checklists for all recommended and required student activities) were developed and provided as a guide to how all elements of the course integrated with the others [e.g., a course blueprint (Coderre, 2009)].

Changes in lecture exam formats throughout the study period included a gradual decrease in the use of publisher provided test banks and an increase in use of self-authored exam questions. Attending professional development sessions on authoring and analyzing exam questions via item analysis at the 2016 HAPS Annual conference (Burgoon and Quinn, 2016), enabled increased use of Learning Outcome linked exam questions and exam validity. The number of questions per exam was reduced from 50 to 45 in the fall of 2019 to assess how the reduction might affect student stress levels and outcomes on the exams. Laboratory practical formats included spotter exams with word banks provided beginning in the fall semester of 2017 (Britson, 2020). Grading equity sessions to standardize the marking of full, partial, or zero credit on lab practical questions (Winter, 2002) began in the fall of 2015 to ensure that student scores were equivalent from one TA to another.

No-risk, cumulative bonus questions (typically 8-10 questions) were added to the last exam in the course as a means for students to increase their performance in course. Students would benefit if questions were answered correctly but would not lose points if answered incorrectly. For the purposes of this manuscript, bonus question points or extra credit points are not included in the earned course scores presented later. Weighting of assessments used to calculate the earned course scores across the study period are presented in Table 2.

In the fall semester of 2020, most courses at the University of Mississippi were delivered via the internet. In Human A&P I, lecture content was pre-recorded and available through the university’s Learning Management System (BlackBoard™). The lecture time period (i.e., 8-8:50am, MWF) was used for “Q&A” sessions, via video conferencing (Zoom™), in which students could ask for additional examples, clarifications, explanations, etc., of any course content. A laboratory teaching assistant was available to help moderate the chat window, admit students to the session, and aid the instructor. Students were required to log in to at least one Q&A session per week. Lecture Q&A sessions were not held on days in which students had lecture exams or laboratory practicals.

Additionally, the Fall 2020 semester was shortened by 1.5 weeks such that final exams were completed before the Thanksgiving holiday. This change in course length necessitated elimination of the basic chemistry content (HAPS Learning Outcomes Modules C1 to C5; HAPS, 2019a) and a reduction in the number of lecture exams from 5 per semester to 4. Except for module removal, the sequence of remaining course content was the same as in prior semesters. Lecture exams were taken through BlackBoard™

<table>
<thead>
<tr>
<th></th>
<th>Fall 2011 - 2015</th>
<th>Fall 2016 - 2019</th>
<th>Fall 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture exams</td>
<td>65%</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>Lecture quizzes (in person)</td>
<td>10%</td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Online homework</td>
<td></td>
<td>10%</td>
<td>18%</td>
</tr>
<tr>
<td>Laboratory quizzes</td>
<td>8%</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>Laboratory formative assessments</td>
<td>7%</td>
<td>7%</td>
<td>6%</td>
</tr>
<tr>
<td>Laboratory practicals</td>
<td>10%</td>
<td>10%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 2. Assessment weightings used to calculate course grade in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020.
with remote video monitoring (e.g., Proctorio™; Woldeab and Brothen, 2021). The number of questions and length of each exam remained at 45 questions and 50 minutes. On the day of the exam, students were able to complete the exam between 8am and 5pm. Questions were randomly selected from a curated pool of questions, per content chapter, and randomly presented all at once to students once they completed the Proctorio™ verification process.

Lab sessions were held synchronously (1-hour 50 min time periods, once per week) via Zoom™ and were led by 2 laboratory teaching assistants (TA). As in previous semesters, sessions began with a pre-lab quiz followed by a short introduction of lab objectives by the TAs. Students were then randomly assigned to groups to 3-4 students to complete the in-lab formative assessment. In addition to online publisher content (e.g., Mastering A&P™ Pearson Education, 2021a; Practice Anatomy Lab™, Heisler, et al., 2013; etc.), resources such as Lt Online™ (ADInstruments, 2022) allowed us to continue use of PowerLab activities similar to the in-person lab experience. Unlike previous semesters, lab sessions began the first week of classes allowing us to include all lab topics and to meet with students regularly. The two lab practicals tested students over the same topics and to meet with students regularly. The two lab practicals tested students over the same topics and used photographs of the models, histological specimens, dissection specimens, etc. Lab practicals consisted of 50, free-response identification question with word banks taken through BlackBoard™ with remote video monitoring. Students were allowed 60 minutes to complete the practicals. TAs were required to confer with the instructor on issues of partial credit, no credit, etc. for grade normalization.

Raw scores for lab practicals from the Fall 2020 semester were normalized to an average of 75 using a z-score transformation (Winter, 2002) due to particularly low scores prior to their use in calculating the student’s earned score for the course.

For analyses, year was considered the independent variable with the following dependent response variables: highest ACT, high school GPA, undergraduate GPA, earned course score (no extra credit included), exam average, lecture quiz average, lab quiz average, lab practical average, average performance on online homework, number of SI session attended, number of missed lectures, number of missed labs, number of missed lecture assignments, and number of missed online assignments. Assessment averages were only calculated with scores from completed assessments. Assessments not attempted were tabulated by assessment category. No additional independent variables were used as doing so would force use of an experimental design where none existed. Rather, analyses were conducted to explore relationships between variables and patterns over time.

A one-way analysis of variance (ANOVA) was used to analyze temporal changes in the dependent variables. Bivariate correlation analyses were performed to explore relationships between the independent variable and dependent variables. Effect size (Cohen, 1988; Lakens, 2013) was calculated for all ANOVA (partial Eta squared) and correlation (Pearson’s R) analyses. Due to the large sample size, the level of significance was set at p=0.001 for both types of analyses. Since the large number of error degrees of freedom in the analytical models makes significant differences more likely as well as Type I errors of interpretation, G*Power 3.1 was used to calculate the minimum sample size required to achieve significant results with a large effect size at the p=0.001 level for the ANOVA (Faul et al., 2007) analyses. Randomly selected subsets of the data set were then used to conduct these conservative analyses. A Discriminant Function Analysis, with letter grade earned used as the grouping variable, was performed to determine which variables contribute the most to variation in the data set and how these variables discriminate between the groups. All analyses were performed using SPSS (Statistical Package for the Social Sciences), Version 27, licensed to the University of Mississippi.

Results

Statistically significant differences were found between ACT scores (F = 29.931; df = 9,3520; p < 0.001), high school GPA (F = 25.301; df = 9,3508; p < 0.001), and undergraduate GPA (F = 20.704; df = 9,3617; p < 0.001) for the initial enrollment of students in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020 (Table 3). Effect sizes (partial Eta-squared, h²) were medium for both ACT (h² = 0.071) and high school GPA scores (h² = 0.061), and small for undergraduate GPAs (h² = 0.049). A random selection of 210 student records (5.7%; rounded to 6% for SPSS operations which allowed only whole numbers) was obtained to meet the minimum sample sizes required for detecting significant results with a large effect size at p = 0.001. There were no significant differences in ACT scores (F = 1.596; df = 9,175; p = 0.12), high school GPA (F = 0.815; df = 9,178; p = 0.603), or undergraduate GPA (F = 1.018; df = 9,181; p = 0.428) from the random selection of 6% of students initially enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020. Sample sizes will vary for reasons listed earlier (e.g., students may not have submitted a score, may have entered as a transfer, or may have just started their undergraduate education).
Table 3. Prior performance scores for students at the time of initial enrollment in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020. Within each categorical variable, means with the same letter are not significantly different at the p = 0.001 level. Sample sizes are not equal across years because the students may not have submitted a score, may have entered as a transfer, or may have just started their undergraduate education.
Among the dependent response variables (Table 4), the number of missed lecture assignments, number of missed lab assignments, and number of missed lectures were not significantly different as calculated from a one-way, univariate ANOVA with either the overall data set or a random subset of the data for testing with the minimum sample size to detect significant differences at the p=0.001 level. Performance on lecture quizzes, pre-lab quizzes, and in-lab formative assessments were significantly different at the p=0.001 level using the overall data set, though the effect size was small for each.

<table>
<thead>
<tr>
<th></th>
<th>data set</th>
<th>random selection</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>h²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>earned course score</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3295</td>
<td>10.349</td>
<td>&lt;0.001</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,217</td>
<td>3.032</td>
<td>0.002</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td><strong>exam average</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3295</td>
<td>15.605</td>
<td>&lt;0.001</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,217</td>
<td>2.648</td>
<td>0.006</td>
<td>0.099</td>
<td></td>
</tr>
<tr>
<td><strong>lecture quiz average</strong></td>
<td>2011-2019</td>
<td>na</td>
<td>8,3040</td>
<td>15.718</td>
<td>&lt;0.001</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>8,199</td>
<td>2.086</td>
<td>0.039</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td><strong>average performance on online homework</strong></td>
<td>2016-2020</td>
<td>na</td>
<td>4,1612</td>
<td>53.845</td>
<td>&lt;0.001</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>21% (p=0.001)</td>
<td>4,355</td>
<td>21.179</td>
<td>&lt;0.001</td>
<td>0.193</td>
<td></td>
</tr>
<tr>
<td><strong>pre-lab quiz average</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3052</td>
<td>19.414</td>
<td>&lt;0.001</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,203</td>
<td>2.164</td>
<td>0.026</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td><strong>in lab formative assessment average</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3294</td>
<td>6.631</td>
<td>&lt;0.001</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,217</td>
<td>0.797</td>
<td>0.619</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td><strong>lab practical average</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3293</td>
<td>26.944</td>
<td>&lt;0.001</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,217</td>
<td>3.681</td>
<td>&lt;0.001</td>
<td>0.132</td>
<td></td>
</tr>
<tr>
<td><strong>number of SI sessions attended</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3295</td>
<td>55.971</td>
<td>&lt;0.001</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,217</td>
<td>5.709</td>
<td>&lt;0.001</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td><strong>number of missed lectures</strong></td>
<td>2018-2019</td>
<td>na</td>
<td>1,667</td>
<td>5.68</td>
<td>0.017</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>50% (p=0.001)</td>
<td>1,331</td>
<td>1.793</td>
<td>0.182</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td><strong>number of missed lecture assignments</strong></td>
<td>2011-2019</td>
<td>na</td>
<td>8,3053</td>
<td>0.751</td>
<td>0.646</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>8,199</td>
<td>0.533</td>
<td>0.831</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td><strong>number of missed lab assignments</strong></td>
<td>2011-2020</td>
<td>na</td>
<td>9,3295</td>
<td>1.021</td>
<td>0.42</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>7% (p=0.001)</td>
<td>9,217</td>
<td>0.351</td>
<td>0.956</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td><strong>number of missed online assignments</strong></td>
<td>2016-2020</td>
<td>na</td>
<td>4,1612</td>
<td>15.049</td>
<td>&lt;0.001</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>21% (p=0.001)</td>
<td>4,355</td>
<td>7.751</td>
<td>&lt;0.001</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.* One-way, univariate analysis of variance (ANOVA) results with effect sizes for selected response variables from for students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020. Overall sample size may differ due to inclusion of a resource (e.g., online homework or tracking of lecture attendance) or the COVID-19 pandemic altering course formats. Random samples were selected to analyze the minimum sample size to detect significant differences at the p=0.001 level.
The earned course score (Fig. 1), exam average (Fig. 2), and number of missed online assignments (Fig. 3) were significantly different at the p=0.001 level using both the overall data set and the minimum sample size required to detect significant differences at that level. For each variable the effect was small with the overall sample size and medium with minimum sample size. Student-Newman-Keuls post-hoc pairwise comparison tests using the overall sample size revealed that the earned course score was significantly higher for students in 2020 as compared to students from all other semesters, and the number of missed online assignments was significantly lower for 2020 students as compared to students from all other semesters. The exam average was significantly higher for students from 2013, 2014, and 2020 as compared to students from all other semesters.

**Figure 1.** Mean course score (+1 standard deviation) earned by students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 through 2020. Means with the same letter are not significantly different at p < 0.001. Yearly sample sizes are depicted by dots according to the secondary axis.

**Figure 2.** Mean exam average (+1 standard deviation) earned by students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 through 2020. Means with the same letter are not significantly different at p < 0.001. Yearly sample sizes are depicted by dots according to the secondary axis.

**Figure 3.** Mean number of missed online assignments (+1 standard deviation) earned by students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2016 through 2020. Means with the same letter are not significantly different at p < 0.001. Yearly sample sizes are depicted by dots according to the secondary axis.
The average lab practical score (Fig. 4) was significantly different at the p=0.001 level using both the overall data set and the minimum sample size required to detect significant differences at the p=0.001 level with a medium effect size for each. Student-Newman-Keuls post-hoc pairwise comparison tests using the overall sample size revealed that the average lab practical score was significantly lower for students in 2015 and 2020 as compared to students from all other semesters.

Both the average score on online assignments (Fig. 5) and number of SI sessions attended (Fig. 6) were significantly different at the p=0.001 level using the overall data set and the minimum sample sizes required to detect significant differences at the p=0.001 level. There was a medium effect size on the average score on online assignments for the overall sample size and the minimal sample size for testing at p=0.001. Student-Newman-Keuls post-hoc pairwise comparison tests using the overall sample size revealed that the average score on online assignments increased from 2016 to 2020. There was a medium effect size for the number of SI sessions attended using the overall sample size and a large effect when using the minimal sample sizes needed for testing at the 0.001 level. Student-Newman-Keuls post-hoc pairwise comparison tests using the overall sample size revealed that fewer SI sessions were attended in 2011, 2012, 2013, and 2020 as compared to all other semesters.

Figure 4. Mean lab practical average (+1 standard deviation) earned by students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 through 2020. Means with the same letter are not significantly different at p ≤ 0.001. Yearly sample sizes are depicted by dots according to the secondary axis.

Figure 5. Average performance on online homework (+1 standard deviation) earned by students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2016 through 2020. Means with the same letter are not significantly different at p ≤ 0.001. Yearly sample sizes are depicted by dots according to the secondary axis.

Figure 6. Mean number of supplemental instruction (SI) sessions (+1 standard deviation) attended by students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 through 2020. Means with the same letter are not significantly different at p ≤ 0.001. Yearly sample sizes are depicted by dots according to the secondary axis.
Significant, pairwise correlations (two-tailed) between the independent variable (semester) and dependent variables were numerous (Table 5). Of the 74 possible pairwise comparisons, only three (semester vs. number of missed lectures, number of missed lecture assignments, or lecture quiz average) were not significant. Eight of the 74 possible pairwise comparisons involved the independent variable (semester), but the effect size was small for 7 of these comparisons, with the remaining comparison at a medium effect size. Of the remaining pairwise comparisons, 51 had a large effect size exceeding $r = 0.5$.

| Table 5. Pearson correlation coefficients ($r$) for students enrolled in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020 (maximum sample size, $n=3305$). Correlation coefficients that exceed ±0.5 (i.e., a large effect size) for a significant ($p<0.001$) correlation are underlined, and correlations on the diagonal are not depicted.
In the discriminant function analysis using the overall data set (Fig. 7), the first discriminant function (DF1) explains 89.5% of the variation in the data set with an Eigenvalue of 7.204, and the second discriminant function explains 8.1% of the variation with an Eigenvalue of 0.648. Using this model, 83.6% of the cases are classified correctly into the grouping variable (e.g., letter grade). Variables loading most heavily onto DF1 are the exam average (0.849), lab practical average (0.609), and lab quiz average (0.414); all variables expressing student performance. Variables loading most heavily onto DF2 are the number of missed lecture assignments (0.735), number of missed lab assignments (0.602), and number of missed online assignments (0.374); all variables expressing student behavior (i.e., engagement). The only engagement variable loading positively to DF1 was the number of SI sessions attended (0.143). The number of SI sessions attended is significantly correlated with the exam average (Fig. 8) using the overall data set, although the effect size is small (r=0.198).

**Figure 7.** Discriminant function analysis plot grouped by letter grade recorded for students enrolled in Human Anatomy and Physiology I at the University of Mississippi (n=3304) from 2011 through 2020. Each data point represents the paired discriminant functions for the combined loadings for the variables: exam average, lab practical average, lab quiz average, missed lecture assignments, missed online assignments, percent correct of attempted online homework, in lab formative assessment average, number of missed lab assignments, lecture quiz average, missed lectures, and number of Supplemental Assessment sessions attended. Variables that load the most to Discriminant Functions 1 (Performance) and 2 (Engagement) are mentioned on the x and y axes. In the legend, classification results of the discriminant function analysis are listed as percent correctness for each grouping variable.

**Figure 8.** Correlation between the number of Supplemental Instruction (SI) sessions attended and average exam score, with regression line, earned by students enrolled in Human Anatomy and Physiology I at the University of Mississippi (n=3304) from 2011 through 2020.

**Discussion**

When statistically significant results from a study are numerous, it is no easier to gain insight into effective teaching practices than when significant results are absent. Including the additional consideration of a large data set increases statistical power but interpretations and conclusions of results would be suspect without measures to evaluate effect size and influence of sample size. Incorporating results from one-way ANOVAs (including those from a randomly selected sample), correlation analyses, and discriminant function classification analyses to identify large and persistent patterns within a large data set will move us forward in identifying a suite of best, effective practices as well as equitable practices leading to student success in Human Anatomy and Physiology.

Students come to Human A&P I at the University of Mississippi from a diverse background in terms of both academic preparation and career goals. The omission of linked preparation scores to course scores in Human A&P I is an unfortunate reality that must be acknowledged in presenting interpretation of these results. Going forward with yearly inputs to the data set, permission will be obtained from students enrolled in Human A&P I to link their preparation scores with course performance scores. I recommend that all instructors who track long term data in their courses begin obtaining permission to link scores to increase the robustness and completeness of future analyses.
Significant differences, and steady increases, were found throughout the preparation score data set (entire sample) for highest ACT score, high school GPA, and undergraduate GPA. However, the effect sizes were small to medium, and the difference between the minimum and maximum mean ACT score, high school GPA, and undergraduate GPA were 3.64 points, 0.84 points, and 0.4 points, respectively. Minimum levels of each variable are important for admittance and scholarship offers at the University of Mississippi (U Miss, 2022) as well as many other institutions. But do these variables add information that will help us as instructors identify the skills and background information that students bring to our courses enabling us to be more effective at teaching the skills and content that our students need in our courses? Statistically significant differences between ACT scores, high school GPA, and undergraduate GPA were absent when analyzing a randomly selected, minimum sample size needed to identify significant results with a large effect size at the p=0.001 level.

In light of the overall lack of significance with minimal sample sizes, subsequent results from analyses on engagement and performance can be interpreted with reduced, though not eliminated, consideration of the potentially confounding variables of ACT score, high school GPA, and undergraduate GPA. Hopp (2009) found successful completion of a prior Chemistry course to be an indicator of success in A&P I. Results from prior studies (e.g., Russell et al. 2016; Sturges et al., 2016; Young 2019; Keller and Hughes, 2021), indicate that prior content knowledge and/or academic skills are correlated with success in Human A&P I. Should a specific level of knowledge or skills be consistently identified from multiple investigators across institutions, how do we and should we standardize quantification of this knowledge/skill as an entry tool for enrolling in Human A&P? Conversely, as more institutions offer test-optional admissions as a means to increase equitable access to higher education (Lewis et al., 2021), would we be cancelling out increased access by enforcing minimum standards? Use of the HAPS Human A&P I Exam (HAPS 2019c) as a standardized entry tool would be helpful for knowing what content knowledge our students bring to our courses, allowing us to modify course delivery (to meet the Learning Objectives) rather than prevent enrollment. However, the HAPS A&P Exam does not give us an indication of the study skills, self-motivation, and self-discipline that are also critical for success in Human A&P I. A studying examining the role of prior skills, entry level knowledge, use of standardized MSLQs (Orsini et al., 2016; Abdel-Meguid et al., 2019), and student outcomes is needed to move forward with continued improvement and implementation of teaching practices.

Qualities of engagement in the A&P student include regular interaction with course content, making connections between content areas, preparation for class or lab activities, taking advantage of study sessions or open lab sessions, willingness to help classmates, focus on coursework, willingness to ask questions, managing course responsibilities, etc. (Johnson and Gallagher, 2021). These are all qualities that instructors promote with a curriculum of effective teaching practices such as course blueprinting, backward design, high structure, and retrieval practice. Using the adjective of “effective” for teaching practices implies objective quantification of these qualities through performance on course assessments. However, these teaching practices help all students in all areas of an A&P course rather than just promote engagement, and isolating increases in engagement from performance can be difficult.

Attendance in lectures, labs, and study/SI sessions as well as completion of assignments is a simple measure of engagement because it is assumed that increases will be seen in all of these variables in the engaged student (Dogrell, 2021). In the present study, isolation of these variables from performance enables use of a more objective assessment variable for engagement, and possible one that could be incorporated into a course dashboard. Students are generally reliable in their attendance in lectures, labs, submission of in-lecture assignments, and submission of in-lab assignments as there were no differences in these variables across the years of the study in the entire data set or subsets. Similar to the report of Utz and Bernacki (2018), it is in the voluntary attendance (e.g., SI sessions) and submission of frequent, low-stakes assessments (e.g., online homework) that differences are detected. Attendance at SI sessions increased once a small incentive was offered to students attending. While attendance at any single SI session would only add a maximum of 0.08% extra credit points to a student’s overall course percentage, each SI session attended was correlated with a 0.739% increase in their exam average. This increase in exam average would lead to an increase of 0.4% to the overall course percentage, for each SI session attended. The gain in performance is greater than the incentive. The number of missed online assignments decreased over time with only the 2020 mean number being significantly lower than the other years in which online assignments were required. This difference will be discussed later in the pandemic teaching section of the Discussion.

Given the number of pedagogically effective course modifications that were implemented during the years of this study, few modifications were found to be directly related to changes in dependent variables once procedures for evaluating effect size and influence of sample size were examined. Meaningful changes (e.g., backwards design, course blueprinting, high-structure, retrieval practice, decrease in TA to student ratio in the lab, selection of different publisher resources, questions per exam, etc.) were made with a great deal of research and deliberation prior to implementation, but student performance remained stable. Many of these variables were likely interacting with each other resulting in further difficulty in interpreting continued on next page
Ten Years in the Human Anatomy and Physiology I Classroom: A Retrospective Analysis of Student Preparation, Engagement, Performance, and the Impact of COVID-19

results. From a different perspective, the lack of statistically significant results with large effects sizes (e.g., 45 vs. 50 questions per lecture exam; provision of word banks for lab practicals (Britson, 2020)) while still achieving learning outcomes could be considered a positive outcome. Each of these changes were subjectively associated with reduced confusion, stress, and test anxiety. Subjective feedback from SI leaders and TAs indicates that students struggle with a great deal of anxiety associated with their performance goals for the course and potentially unmet expectations (Leary and Bryner, 2021). Ideally, as educators we would like to see simultaneous increases in performance and engagement with decreases in anxiety and stress, and further investigation into the sources, effects, and mitigation of stress in A&P students (Fornier et al., 2017; El-Baze et al., 2018) is warranted.

While the number of significant pairwise correlations between variables were numerous, the correlations show a common theme: both engaged, high-performing students and less engaged, low-performing students are consistent in these attributes within and across the years of the study. In constructing a mathematical model (e.g., a discriminant function analysis) to explain variation within the data set, using letter grades earned as the grouping variable when the letter grade is determined by performance, uses circular logic. However, in the discriminant function analysis performance explains 89.5% of the variation rather than 100%. If performance was the only factor influencing the letter grade, the plot in Fig. 7 would be a horizontal line.

Consistent student attributes of performance and engagement, and the similar percent of passing (A, B, and C) and not passing (D, F, and W) letter grades in the present study as compared to published percentages from Human A&P I courses (over 12,000 students total) across a range of institutions from 2-year colleges to R1 doctoral granting universities (Table 6), indicate that a “A&P Success Dashboard” would be both feasible and useful if timely interventions were identified by instructors and implemented by students. In immediate practice, one of two approaches is used to aid the struggling student. First, if the student is completing and submitting all coursework, the student needs help understanding, and demonstrating that understanding, of course content. Second, if the student is not completing and submitting coursework, the student needs help with time management, self-discipline, and time-on-task focus. Isolating these messages with examination of fixed vs. growth mindsets in students (Stuart and Wolcott, 2021) would be useful data that could be added to a success dashboard.

On initial examination it would appear that the many challenges faced (e.g., internet and technology inequities, lack of engagement, difficulty focusing, time management inconsistencies, etc.; Schaeffer 2022) during pandemic teaching did not translate into decreased performance as

<table>
<thead>
<tr>
<th></th>
<th>Percent Passing (A, B, C)</th>
<th>Percent Not Passing (D, F, W)</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdullahi and Gannon, 2012</td>
<td>36</td>
<td>64</td>
<td>1429</td>
</tr>
<tr>
<td>Atamturktur et al., 2015</td>
<td>59</td>
<td>41</td>
<td>1042</td>
</tr>
<tr>
<td>Beeber and Biermann, 2007</td>
<td>65.4</td>
<td>34.6</td>
<td></td>
</tr>
<tr>
<td>Caplan, 2015</td>
<td>67</td>
<td>33</td>
<td>61</td>
</tr>
<tr>
<td>Griff and Matter, 2008</td>
<td>72</td>
<td>29</td>
<td>222</td>
</tr>
<tr>
<td>Guiltice et al., 2015</td>
<td>67.3</td>
<td>32.7</td>
<td>2728</td>
</tr>
<tr>
<td>Harris et al., 2004</td>
<td>65</td>
<td>35</td>
<td>91</td>
</tr>
<tr>
<td>Hopp, 2009</td>
<td>56.4</td>
<td>43.6</td>
<td>546</td>
</tr>
<tr>
<td>Hopper, 2011</td>
<td>38</td>
<td>62</td>
<td>101</td>
</tr>
<tr>
<td>Keller and Hughes, 2021</td>
<td>69</td>
<td>31</td>
<td>306</td>
</tr>
<tr>
<td>Langtree, 2014</td>
<td>65.7</td>
<td>34.3</td>
<td>248</td>
</tr>
<tr>
<td>Rosenzweig, 2006</td>
<td>51.1</td>
<td>48.9</td>
<td>2236</td>
</tr>
<tr>
<td>Russell et al., 2017</td>
<td>58.4</td>
<td>41.6</td>
<td>3693</td>
</tr>
<tr>
<td>Young et al., 2019</td>
<td>73.3</td>
<td>26.6</td>
<td>60</td>
</tr>
<tr>
<td><strong>Mean, Published Reports</strong></td>
<td><strong>60.25</strong></td>
<td><strong>39.75</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Present Study</strong></td>
<td><strong>61.55</strong></td>
<td><strong>38.39</strong></td>
<td><strong>3305</strong></td>
</tr>
</tbody>
</table>

Table 6. Percent of passing (A, B, and C) and not passing (D, F, and W) grades earned in Human Anatomy and Physiology I at the University of Mississippi from 2011 to 2020 compared to published reports. Percentages for specific letter grades were not available from each report. Successfully passing Human A&P I with a letter grade of C or better is a requirement for enrollment in Human A&P II at the University of Mississippi and is aligned with the curriculum guidelines recommended by the Human Anatomy and Physiology Society (HAPS 2019b).
course scores were higher in fall 2020 than all other years. Confounding this result is that both lab practical scores in fall 2020 were normalized to an average of 75 using z-score transformation, and that there were 4 lecture exams rather than 5 during the semester. The lecture exam average was significantly higher in fall 2020 (70.163%) than all other semesters except for 2013 (67.406%) and 2014 (65.735%). During 2011 to 2019 exams were administered in-person, at 8am, in a large lecture hall seating a maximum of 400 individuals, and in a single class session. In 2020, students had the same 50 minutes to complete an exam but could take the exam at their convenience between 8am and 5pm on the exam day and in their preferred location and testing environment. Students also reported that they had more time to study because there were no social activities competing for their time. Though measures were taken to reduce the potential for academic dishonesty (e.g., remote proctoring, randomized selection and presentation from a question pool, etc.), it is unlikely that all forms of academic dishonesty were eliminated. The number of missed online assignments in 2020 was half of the number observed in the four previous years, potentially due to the lack of distinction between types of assignments as all assignments were submitted remotely. Collectively, a higher exam average, transformed lab practical scores, and fewer missed online assignments likely increased the calculated course score for fall 2020. With all lectures, labs, and testing for Human A&P I returning to the in-person format at the University of Mississippi in fall 2021, further analyses comparing changes in time management, study, hands-on, and testing skills from the pre-pandemic to late-pandemic time periods will provide information on how we teach and interact with our students.

Conclusions

Moving forward with the insights gained from this retrospective analysis will internally include: (1) construction of a performance dashboard within the learning management system that emphasizes engagement as well as performance, (2) identification of students’ intrinsic and external motivations that can then be used to develop engagement strategies that build upon these motivations, (3) identification and mitigation of sources of anxiety in students that affect engagement and performance, and (4) development of the “new normal” of the A&P experience post-pandemic. Externally, I offer these analyses and insights for colleagues searching for a comparison group for their own investigations into student performance.

About the Author

Carol A. Britson, Ph.D., is an Instructional Professor and Associate Chair for Undergraduate Studies in the Department of Biology at the University of Mississippi. She teaches Human Anatomy; Histology; Physiology; and Human Anatomy & Physiology I and II. She also serves as the Secretary (2021-2023) for the Human Anatomy and Physiology Society (HAPS).

Acknowledgments

I thank department chairs, Brice Noonan, Gregg Roman and Paul Lago, and the Department of Biology at the University of Mississippi for their support of the anatomy and physiology curriculum and laboratory resources. I thank Nancy Wiggers with the University of Mississippi Center for Excellence in Teaching and Learning for compilation of Supplemental Instruction attendance records. Lastly, I thank my teaching assistants and SI leaders over the past years without whom none of this retrospective would be possible.

Literature Cited


Ten Years in the Human Anatomy and Physiology I Classroom: A Retrospective Analysis of Student Preparation, Engagement, Performance, and the Impact of COVID-19


continued on next page


Langtree EM. 2014. Factors Contributing to Success in Anatomy and Physiology in First Year Students in the KZNCH Nursing Programme (M.Tech.), Durban University of Technology (Doctoral dissertation). https://doi.org/10.51415/10321%2F1254


continued on next page


---


---


---
Confidential Peer-Evaluation as a Method of Learning in Online University Courses

Simon Lemaire, PhD¹, Gladys Bruyninx, MD², Miriam Grenon, CL³, Madisson Kelleher-Radey, BScN⁴, Alexander Yeuchyk, PhD⁵

¹Department of Cellular & Molecular Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5, slemaire@uottawa.ca (corresponding author).
²Faculty of Medicine, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5, gbruy042@uottawa.ca
³Faculty of Law, University of Ottawa, 57 Louis-Pasteur Private, Ottawa, ON, K1N 6N5, mgren012@uottawa.ca
⁴Faculty of Health Sciences, University of Ottawa, 451 Smyth Road, Ottawa, ON, K1H 8M5, mkell123@uottawa.ca
⁵Teaching and Learning Support Service, University of Ottawa, 136 Jean-Jacques Lussier Road, Ottawa, ON, K1N 6N5, ayeuchyk@uottawa.ca

Abstract

Online teaching of university courses to a large number of students is a major challenge in both its execution and evaluation. The proposed model for solving the logistical issues of these courses is based on using the Wiki tool in the Blackboard learning management system (LMS) to facilitate the submission and evaluation of student essays. The Test tool also permits the collection of confidential peer evaluation based on criteria established by the professor. 220 students studying Human Anatomy and Physiology (A&P) were divided into random groups of 5 by Blackboard to undertake and submit a dissertation on one of 5 different topics and, secondly, to evaluate the work of their teammates. The assessments were collected, calculated, and validated by the professor and teaching assistant in Microsoft Excel and submitted to the Blackboard Grade Center. A survey among students after the course showed a high degree of satisfaction with the criteria related to the completion of the dissertation and peer review assignment, as a method of learning. https://doi.org/10.21692/haps.2022.011

Key words: online teaching, written essays, peer review, learning, anatomy and physiology

Introduction

One of the biggest challenges in teaching a large number of students online or in university classrooms is finding a method to monitor and appropriately evaluate student performance and degree of knowledge and understanding on the subject matter while ensuring that the method does not permit plagiarism or biased results (Distler 2015). The most popular summative assessment method is based on exams with multiple choice questions (MCQs). MCQs are also used by many professors to provide formative self-testing for students in between the summative exams (Lull et al. 2016). This method has advantages in that it makes it possible to quickly collect the answers and determine the marks of large numbers of students using computer grading. MCQ-based exams can be easily proctored in the classroom; however, they are more challenging to proctor when exams are administered online. An important consideration with MCQ-based exams is the fact that the correct answers exist in the lists of answer choices provided in the questions themselves; the students simply have to identify them. This means that, in order to answer MCQs correctly, students can either study hard and thus know and understand the material of the course (McConnell et al. 2015) or, at times, proceed by deduction or guesswork to arrive at the correct answer for some questions. Given the difficulties in assessing the performance of the growing number of university students and the limited number of faculty (Burgess et al. 2014) as well as the emergence of online teaching (Distler 2015), there is a need to develop a method that would allow for assessment of student competency and degree of knowledge while encouraging their engagement, motivation, and accurate understanding of course content. Writing assignments, coupled with peer review of these assignments, appear to be a means of providing quality learning opportunities if properly structured and monitored (Pond et al. 1995).

Peer evaluation has previously been studied in undergraduate courses in pharmacy (Dochy et al. 2006; Storjohann et al. 2019) and biology/chemistry (Finkenstaedt-Quinn et al. 2017; Shultz and Gere 2015). These studies assessed the implementation of peer- and self-grading systems and showcased positive results. The methods relied on the development of a great degree of honesty, content knowledge, and self-confidence by the students and they appeared to provide a valuable learning experience for students when coupled with effective instructor guidance.
In this study, students in a large human anatomy and physiology (A&P) class were divided into random groups of five and asked to complete a two-part assignment in which they each wrote an essay on one of five assigned topics and then conducted confidential peer evaluations of the four other written assignments completed by their group of peers.

Methods

General procedure

Within the class of 220 students, random groups of 5 students were formed using the Blackboard LMS. A content area in Blackboard was also created to define the “Rules of the Dissertation” (Table 1) which detailed the subject topic options, the content, the procedure, and deadlines for the dissertation. At the beginning of the course, an oral presentation of 15-20 min was given to explain the “Rules of the Dissertation” and emphasize the various tasks to be completed.

The steps were as follows:

1. Each student in the group first chose one of the five topics related to the different body systems studied in A&P: pernicious anemia (digestive system), hypothyroidism (metabolism), acquired immunodeficiency syndrome (immune system), renal failure (acute or chronic; renal system), or infertility (male or female; reproductive system). Students then published both their subject choice and basic references to be used for their dissertation in the Wiki group. The references published in Wiki allowed the professor or teacher’s assistant (TA) to have rapid access to the information. Subject choices were first come first served as students in the same group could not choose the same topic.

2. Approval of student selected topic and references was done by the professor or TA by making a brief comment on the student’s Wiki page to endorse the literature or make suggestions.

3. Following subject approval, students composed their essay by adhering to the directives outlined in the content area entitled “Rules of the Dissertation.” Briefly, the dissertations were to contain the following elements: a brief summary of the subject (maximum of 200 words), a description of signs and symptoms and possible cause(s) of the disease, a description of the organs and/or systems and/or mechanism involved, a description of treatment and prevention methods of the disease, and references (maximum of 10-12). This work with a limited number of single-spaced pages (2-3 pages) could be supplemented with figures and tables with appropriate legends. Student submitted their dissertation on the Wiki page within their group.

4. Each student read and evaluated the dissertations of their teammates according to the following criteria: a) clarity; b) coverage of the subject and c) overall quality and relevance of the work. They submitted their evaluations in percentages, using a confidential assessment grid (Table 2) through the Test tool in Blackboard. The Test tool allowed the professor (or TA) to export students’ grades and comments to Excel, evaluate the variability and reliability of the grades, and relay the feedback to the students.

Dissertation target dates

The dissertation assignment had three deadlines namely for (1) the choice of the subject, (2) the submission of the dissertation, and, (3) the submission of the peer evaluations (Table 1). Evaluation criteria were established and listed in the “Rules of the Dissertation” as well as in an “Evaluation Grid” used to collect the peer evaluation data (Table 2). Throughout the semester, the professor (or TA) used Blackboard to verify that students were completing the tasks on time and to identify those students who needed to complete one or more given tasks. In the latter case, late students were notified.

The marks and comments entered in the “Evaluation Grid” by the students were solely viewed by the professor or the TA in the Blackboard Grade Center from which the data were downloaded in the form of an Excel spreadsheet. Working within Excel, the professor (or TA) calculated the averages, the standard deviations, and variances of the marks to verify and detect any evaluations that fell outside the range. The final grade for each student was calculated in Excel by the professor (or TA) and uploaded in a new column in the Blackboard Grade Center to be viewed by students.

To conserve confidentiality, student comments were randomly number coded from 1 to 220 and displayed in a spreadsheet on the Blackboard announcement page. The codes of the students were uploaded in a new column of the Blackboard Grade Center so that each student could find their individual code number (and feedback from peers on their essay). The professor (and/or TA) also graded some students who either received a nonconforming evaluation (high margin of error between the assessments) or were assigned to a smaller group due to course withdrawal by some students. Students who did not submit their work, or submitted it too late and without justification, or who incorporated plagiarism or provided biased assessments were informed of the possibility of receiving a grade of zero (Table 1).
Confidential Peer-Evaluation as a Method of Learning in Online University Courses

Examples
To facilitate the dissertation process, a content tab was created to show examples of search engines (PubMed, CINAHL, Medline, PasseportSanté, etc.) that the students could use to find appropriate documentation on their subject. An example of a dissertation on a different topic (Crohn disease (digestive system)) was provided to illustrate the format and quality level that was expected from each student. Finally, an evaluation of the provided example was also shared to show how to analyze a dissertation according to the specific criteria established for the essay and how to formulate a sound review to highlight its strengths and weaknesses.

D2L Brightspace learning platform
The method described above with the Blackboard LMS can also be applied to other learning platforms including, for example, D2L Brightspace. The “Group” tool in D2L can be used to perform events 1, 2, 3 and 4 in Table 1, while the collection of confidential student evaluations (event 5) can be accomplished using the D2L “Quiz” tool with the evaluation grid illustrated in Table 2.

<table>
<thead>
<tr>
<th>THEMES</th>
<th>DESCRIPTION</th>
</tr>
</thead>
</table>
| A. TOPICS | 1. Pernicious anemia (digestive system)  
  2. Hypothyroidism (metabolism)  
  3. Acquired immunodeficiency syndrome (immune system)  
  4. Renal failure (renal system)  
  5. Male or female infertility (reproductive system) |
| B. CONTENT | Brief summary; introduction; problem presentation; discussion; references; tables and figures. |
| C. EVENTS | 1. Formation of random groups of students by Blackboard.  
  2. Student choice & registration of a dissertation subject plus select references in a Wiki.  
  3. Approval of the subject and documentation by the professor or TA.  
  4. Writing and submission of the dissertation by the student within the Wiki. |
| D. EVALUATION | Values (%) based on criteria (1. Clarity; 2. Coverage of the subject and 3. Overall quality and/or relevance) are set for the essay. Use of the confidential “Evaluation Grid” for the student submission of teammate evaluation grades and comments (strengths and weaknesses). |
| E. TIME LIMITS | Dates (deadlines) are established for:  
  1. The choice of the subject.  
  2. The submission of the dissertation.  
  3. The submission of the evaluations. |
| F. PENALITIES | Unjustified late submissions, plagiarism and biased evaluations are prone to be downgraded to zero. |

Table 1. Rules of the dissertation.
Confidential Peer-Evaluation as a Method of Learning in Online University Courses

Results

Validation of student evaluations

The work was completed in accordance with the deadlines by all students. The dates of the deadlines were spread over the entire course session to allow the students time to undertake and enjoy each step, all of which are integral parts of the learning process. The essay counted for 10% of the final grade. The student evaluations were collected and averaged in Excel. The variances (VARP) of the marks (4 per student) were calculated and those that exceeded 7% of the average mark were judged as unreliable and the dissertations were re-evaluated by the professor and TA. Out of 220 evaluations, only 9 were judged to be suspicious. However, after re-evaluation by the professor (and TA), only two of them were slightly modified, indicating that the student evaluations were essentially valid.

Evaluations of the dissertations by the students tended to be higher than those verified by the TA (average of 89% as compared with 84% on a sample of 4 groups of students) or obtained with the final MCQ exam (average of 86% as compared with 76% for the whole class of 220 students). One way to avoid overvaluation (or undervaluation) was to warn the students in advance that they could be penalized if they gave their classmates marks that were too high or too low without proper justification in the comments section of the evaluation rubric. On the other hand, the high grading of students in these essays proved to be valid based on the stronger work of the students in the essays compared with the general closed-book examinations, in part due to having free access to the documentation while generating their dissertations.

Student satisfaction and possible use of D2L Brightspace instead of Blackboard LMS

A survey of the students (Table 3) after this exercise demonstrated a high degree of satisfaction for this method of learning and self-evaluation (Figure 1). Students reported satisfaction levels of 86.9%, 84.4%, 56.4% and 63.5% with

<table>
<thead>
<tr>
<th>Name of student</th>
<th>Group number</th>
<th>Clarity (%)</th>
<th>Coverage of topic (%)</th>
<th>Overall quality &amp; significance (%)</th>
<th>Comments to justify evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Rubric used by students to evaluate the dissertations of their 4 teammates.

Figure 1. Student satisfaction rate concerning the clarity, relevance, confidential peer evaluation and online method of learning.
Confidential Peer-Evaluation as a Method of Learning in Online University Courses

regard to the criteria of: clarity of instructions, relevancy of topics, validity of peer evaluation marks, and value of the online method of learning through peer-reviewed dissertations, respectively. Most of the student comments were positive, and favoured the continuation of this method of learning, especially for A&P courses which demand significant memorization in preparation for summative examinations. This method of teaching and learning a specific A&P subject was novel and refreshing for us compared to the traditional methods of teaching and learning. Some comments indicated the importance of properly instructing the students about the necessity to read and understand their classmate's essays prior to evaluating them according to the established criteria. Following the switch from Blackboard to the D2L Brightspace LMS at the University of Ottawa, the method was applied to other groups of students with the same ease, and peer reviews were collected in the same anonymous manner as with the Blackboard LMS.

Discussion

The main advantages of peer-evaluations are the incentives for students to both produce high-quality works and provide accurate assessments of their classmates’ essays which ultimately enhance the quality of the learning process itself (Adachi et al. 2018; Chen 2012). Students who know they will be evaluated by their colleagues are generally more competitive and want to submit quality essays. After taking the time to write and submit their own essay, they are eager to learn what their colleagues have submitted. They can then evaluate the work of their colleagues according to the criteria they themselves have previously followed, and with which they are familiar. We found that students were not prone to misjudging their colleague’s work because they knew they could be penalized for giving an unjustified mark that was found to be either too high or too low. Also, in their assessment, they were asked to add short comments to indicate the strengths and weaknesses of the work, thereby justifying their assessments.

In terms of learning, the students achieved this by gathering information on a given subject, integrating that material, and describing the important features, while following the well-defined guidelines for essay format. Furthermore, assessment of the work of four of their colleagues on different topics added to their learning about other aspects of the course. As reported by others, we found that students tended to give their colleague’s works higher marks in comparison with the professor’s (or TA’s) evaluation (Li et al. 2016; Reinholz 2016; Rudy et al. 2001). On the other hand, the higher grades observed with the peer-review essays in comparison with those obtained in the final MCQ exam may have been due, in part, to the fact that the final exam covered the entire course and was timed and closed book, whereas the dissertation essays covered specific aspects of the course and students were given much more time to research and write their articles. Nonetheless, due to a tendency of the essays to have higher scores than the MCQ exams and the fact that they were targeting only certain aspects of course content, we suggest that such works should be worth between 5% and 15% of the final grade.

An ability to properly evaluate the work of peers is part of the learning process required in several schools of medicine and implies that the students must learn to recognize their strengths and weaknesses in writing, in topic knowledge, and in time management (Chen 2012). The group of students referred to herein were primarily Faculty of Health Sciences students registered in an A&P course given by the Faculty of Medicine of the University of Ottawa. However, this teaching and learning strategy could be applied to any other university group, regardless of their discipline. As part of their academic

<table>
<thead>
<tr>
<th>Give a rating of 0 to 5 (0= strongly disagree, 5= strongly agree)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instructions given in class and on the website about the dissertation were clear.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The topic choices were relevant and appropriate to the course objectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I consider my essay evaluation by my colleagues to be fair and valid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall, I consider the confidential peer-evaluation of dissertations to be a good learning tool.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comments:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Rubric used by the students to rate the dissertation process for its clarity, relevance, validity of peer evaluation and online method of learning.
Confidential Peer-Evaluation as a Method of Learning in Online University Courses

Looking at the data collected, we were able to calculate the averages and variances of the marks. This allowed us to determine the evaluation criteria and reassure the students that their effort would not be in vain. We therefore consider that this peer-review method could be applied in another class (unpublished data).

According to Papinczak et al. (2007), the peer review process may have a negative impact on problem-based learning (PBL) groups among medical students. These authors mentioned that collaboration and collegiality within a PBL group can be decreased in response to negative feedback from colleagues. To prevent this possibility, we formed random groups of students by making use of a Blackboard tool that allowed the formation of groups wherein most students did not know one another. Furthermore, the evaluations and comments submitted using the Blackboard Test tool were confidential and the professor collected them and established averages and variances prior to submitting the anonymous marks and comments to the Blackboard grade center. This ability to process marks and comments confidentially allowed the students to get to know their strengths and weaknesses without developing any resentment towards their teammates.

Students in the Faculty of Health Sciences as well as those in Medicine experience pressure to succeed. Due to previous experiences in education, some students may feel uncomfortable or skeptical about the value and validity of peer evaluations (Balantyne et al. 2002). Typically, their academic abilities have been developed to learn a large array of content-specific information, but they generally do not know yet how to proceed to evaluate essays. In this context, the professor’s role is particularly important to determine the evaluation criteria and reassure the students of their capabilities to pick up essential knowledge from other people’s work and to form their own judgments about the work’s merits. Students can develop these abilities quite rapidly after being trained by first creating their own work and then applying the same criteria when evaluating colleagues’ works (Newbold et al. 1995). The assignment described above required the professor and TA to allocate time to verify the different stages of the learning process, to calculate the averages and variances of the collected marks, and, if necessary, to evaluate the essays themselves if scores looked suspicious. Such verification allowed us to detect one case of plagiarism in a group of 60 students when this technique was reapplied in another class (unpublished data). Therefore, it is very important that the professor and/or TA remain vigilant and proactive in order to ensure the success of this peer-review method of evaluation and self-learning.

We did not explore all of the possibilities that this peer-review learning procedure may have for developing reflective and critical thinking as described by Harasym and colleagues (2008). For example, the 3 or 4 best essays on each research topic within the class of 220 students could have been selected with the help of the Blackboard LMS after the compilation of student evaluations and included as reference content testable in the final exam. This way, the students would get to know what can be considered as the required knowledge on each topic. On the other hand, another way to strengthen the student integration and comprehension of the research topics could also be to form student subgroups among the existing groups wherein each student would be put in contact with a determined number of students of other groups having chosen the same subject. These student subgroups could be requested to hold online meetings to discuss the research subject before writing their own essay. This way, their work would result not only from their own research and way of thinking, but also from the discussion with other students preparing their dissertation on the same topic. A brief report of these online meetings could be put on the subgroup Wiki so that the professor or TA could verify the dates and contents of the meetings and monitor or advise the students whenever necessary. Finally, another way to strengthen this peer-review method of learning could also be to ask the students to submit a revised version of their work after receiving and considering the comments of their teammates for improvement of the essays.

Previous studies have reported that students’ unfamiliarity with concepts or peer review activities of essays can be an obstacle to the learning process (Halim et al. 2018). In this regard, we spent much time establishing and explaining the rules of the dissertation and the time limits of the various tasks to be carried out by the students (Tables 1 and 2).

The degree of clarity of the instructions as indicated in the student survey was relatively high. However, some students still indicated in their comments that the rules of the essay became clear to them only after the TA and/or professor provided a PowerPoint presentation to the class outlining the dissertation procedures, suggesting the importance of person-to-person communication for the disclosure of instructions to students.

The rationale for the project described above of groups of 5 students writing an essay on a disease related to a system in A&P followed by peer evaluation by colleagues who have written an essay on a different topic linked to a different body system stands on the need to develop education tools that promote the ability of students to be self-directed learners, to be creative, and to become more fully immersed in their education. In this assignment, students had to conduct research and form their own opinions on issues that were related to systems studied in the A&P course. This type of assignment encourages students to make links between concepts and physiological and/or disease conditions. The research on these new education avenues is justified at this time as many of our courses may remain hybrid or even fully online as we gradually transition our teaching during the post-pandemic period.
About the Authors

Simon Lemaire, PhD, is a Professor in the Department of Cellular and Molecular Medicine at the University of Ottawa. He teaches A&P and pharmacology to graduate and undergraduate students in the Faculties of Medicine, Health Sciences and Science. His research focuses on developing online and hybrid courses for students. Gladys Bruyninx, BSc, is an undergraduate medical student at the University of Ottawa. She was a TA for the A&P course. Miriam Grenon, BSc (Hons), is a JD 2022 candidate at the University of Ottawa and was also a TA for the A&P course. Madisson Kelleher-Radely, BScN, was a TA for the A&P course and was particularly involved in collecting the anonymous survey data. Alexander Yeuchyk, PhD, was with the Teaching and Learning Support Service for the development of hybrid courses at the University of Ottawa. He is now a professor in the Languages Department, University of Ottawa.

Literature Cited


Online Tutoring System (MoFaCTS) for Anatomy and Physiology: Implementation and Initial Impressions

Amanda M. Banker, MS¹, Philip I. Pavlik Jr, PhD², Andrew Olney, PhD³, Luke G. Eglington, PhD⁴

¹Southwest Tennessee Community College, 737 Union Ave, UNS 118G, Memphis, TN, 38103, ambanker@southwest.tn.edu (corresponding author)
²University of Memphis, 365 Innovation Dr., Suite 303, Memphis, TN 38152, ppavlik@memphis.edu
³University of Memphis, 365 Innovation Dr., Suite 303, Memphis, TN 38152, aolney@memphis.edu
⁴Amplify, 81 Albion St., #3, Medford, MA 02155, luke.eglington.mail@gmail.com

Abstract

The Mobile Fact and Concept Textbook System (MoFaCTS) is an individualized online tutoring system designed to increase information comprehension and retention. It is being implemented in community college anatomy and physiology (A&P) courses for further system development. A&P was selected because it is a very challenging and highly in demand course. MoFaCTS was used to create Cloze (fill-in-the-blank) questions from the course text which students use to practice the course material. The system provides correction and feedback to the student, repeating questions to improve recall. The system also produces detailed progress reports for both faculty and students. A survey given to students showed a moderately positive impression of the systems, with A&P II students responding more positively on some survey items.

https://doi.org/10.21692/haps.2022.012

Key words: Adaptive learning, reading comprehension, memory, mental model, artificial intelligence

Introduction

Anatomy and physiology (A&P) courses are notoriously challenging. They are also required courses for many fields of study. For example, nursing, physical therapy, dietetics, sports sciences, radiological technician, medical assistant, and more require this course as part of the program or as a prerequisite. A&P courses are high enrollment courses with low success rates, a situation which impacts the future academic goals of some students. Many programs require a minimum grade of C in A&P and successful acceptance into a program is necessary for students to continue in their intended field of study. Additionally, students who fail also risk losing financial aid, a consequence that can negatively impact their future academic progress. While there is variability across multiple institutions, an estimate of course success rates (a grade of C or better) ranges between 55 and 65% (Gultice et al. 2015; Harris et al. 2004; Hull et al. 2016). As such, effective new methods to improve reading comprehension and retention, and ultimately course success, could be valuable.

These courses have an impact on the workforce as well. A&P is a key prerequisite for nursing programs. The United States Registered Nurse Workforce Report Card and Shortage Forecast projects an increasing lack of RNs leading to nationwide shortages by 2030, especially in southern and western areas of the country (Juraschek et al. 2019). The Bureau of Labor Statistics expects Registered Nursing (RN) to see huge growth, with 1.09 million job openings by 2024 (Hogan and Roberts 2015). There will be a great need for students who have successfully completed two A&P courses and low A&P completion rates may create a shortage of nurses as demand grows. Methods for improving course success would greatly impact individual students and, more broadly, the allied health workforce. For these reasons, A&P faculty are always searching for ways to improve student success.

The Mobile Fact and Concept Textbook System (MoFaCTS) is an online practice system developed to improve recall and retention of course material with personalized learning tailored to individual student ability and knowledge. It is currently being implemented in community college A&P classes to improve its efficacy, both generally and specifically in this area of study. It is an adaptable system that can use a textbook to generate questions and then present these questions with correction and feedback.

Background

While various reasons for low success rates in A&P have been proposed and evaluated, the most malleable cognitive factors are reading skill and background knowledge in foundational science topics (Harris et al. 2004; Hull et al. 2016). Reading is a perennial problem for college students. Even though 69% of high school graduates attend college, only 37% of high school graduates are proficient readers as defined by the National Assessment of Educational Progress (National Center for Educational Statistics 2015; US Bureau of Labor Statistics 2022). It is no wonder, then, that A&P students struggle with their textbooks, especially when the textbooks
themselves require a high reading level. For example, Hole’s *Human Anatomy & Physiology*, a common A&P text also used at Southwest Tennessee Community College (Southwest), has a Lexile score of 1260L, placing it at the upper end of the 12th-grade reading level. Since only approximately half of first-year college students are proficient readers (Olney et al. 2017; National Center for Educational Statistics 2015; US Bureau of Labor Statistics 2022), content-area reading deficiencies pose a serious problem for postsecondary education.

Content-area reading deficiencies start at an early age. They are often marked by a sudden drop in reading scores, particularly for students from low-income families (Chall and Jacobs 1983), as students transition to increased reliance on textbooks for learning (Hirsch 2003; Moss 2005). These early deficiencies often cause affected students to read less, which delays both fluent reading and vocabulary growth, and negatively impacts reading comprehension (Mol and Bus 2011; Torgesen 2004). To address this problem, researchers have called for reading comprehension practice to be embedded in the learning of content areas (National Institute of Child Health and Human Development 2000; Snow 2002). Aligning reading comprehension practice with content area learning can help students less familiar with the content area vocabulary, background knowledge, and grammatical style, all of which can cause serious comprehension deficits (Cromley and Azevedo 2007; Fang 2006; Kintsch 1998; Laufer 2013; Nagy and Townsend 2012).

Cloze questions, also known as fill-in-the-blank questions, are well established for vocabulary and comprehension practice and assessment (Fang 2006; McKeown 1985; National Institute of Child Health and Human Development 2000). When no answer alternatives are provided, filling in the blanks requires a student to look carefully at the context and draw on background knowledge to construct an answer. While it is possible that students could use shallow strategies, which would reduce deep meaning-based learning from the Cloze questions, we have designed MoFaCTS to personally adapt to the student (described below) in a way that emphasizes deep processing of items rather than rote memorization strategies. We expect this deeper processing will promote both fluency and understanding of the content.

**MoFaCTS System Development**

MoFaCTS was originally developed in 2006-2007 at Carnegie Mellon University (Pavlik Jr. et al. 2007). MoFaCTS was designed to address the lack of scientifically backed learning tools for content that needs repetition. One original context for MoFaCTS development was Chinese language learning which has many words and characters (Pavlik Jr. et al. 2008). Spaced practice is often recommended in such situations and heuristics exist like the Leitner method for flashcards in which frequency of practice is linked to practice success for each card (Leitner 1972).

However, the MoFaCTS system uses a quantitative memory model based on a scientific theory of memory (ACT-R) to estimate learning as the practice progresses. Inferred from this student practice performance, the model estimates optimal decisions about when to practice items for maximal learning (Eglinton and Pavlik Jr 2020; Pavlik Jr and Anderson 2008). While comparisons with all possible heuristic algorithms are not possible, optimal learning (using a student model) is expected to be more effective than heuristic algorithms in most situations. The system’s goal is to speed simple learning tasks, saving the student time and allowing faster acquisition of key facts in a domain.

In Fall 2019, a three-year grant was awarded through the Institution for Education Sciences (IES) to develop the MoFaCTS System, culminating in a pilot study of the system in the 2022 Academic Year. The current report describes our development progress and student attitudes toward the system. It is useful to note that the work described in this paper represents the first two years of our IES project, and the Department of Education mandates that we focus on developing the system for efficacy testing during this period. We look forward to eventually reporting on the fully developed system and the efficacy test.

**A&P at Southwest Tennessee Community College**

Southwest is an open-enrollment, two-year community college that offers associate degrees, technical certificates, courses for transfer, and prerequisite courses for many career programs. The Southwest student body is diverse, consisting of 61% Black/African American, 7% Hispanic, 26% White, and 6% Other (TBR: The College System of Tennessee, 2021a, b). Most Southwest students are female (67%), and many qualify as low income (41%) (TBR: The College System of Tennessee, 2021a, b, c). As with many colleges, students are often academically underprepared for college courses. The average American College Testing (ACT) results of first-time, first-year students is consistently below the threshold and therefore requires learning support coursework, which must be completed before registering for A&P (Office of Institutional Research 2018).

Southwest offers many programs in allied health, including nursing, physical therapy assistant, radiological technician, and funeral services. All of these require at least one semester of A&P. As a component of many programs, A&P courses are in high demand. On average, Southwest offers 65 sections of Human A&P I and 47 sections of Human A&P II per year, although not all participated in this study. The students taking these courses often have difficulty succeeding. For example, Southwest Community College’s internal research found that for the 2017–2018 academic year, A&P I students had a 35% rate of D, F, or W (withdrawal) grades and A&P II students performed only a little better at a 30% D/F/W rate. These educational challenges are not unique to Southwest students. The development of the MoFaCTS system at Southwest reaches a student population with challenges widespread among college students.

continued on next page
**Methods**

**Content Generation**

MoFaCTS creates automatically generated Cloze questions based on text input into the system (Figure 1). Text analysis methods identify key sentences and select words for fill-in-the-blank questions. These methods decide what sentences and words are most important and form the ‘backbone’ of the text, both at the chapter section level and the overall chapter level (Pavlik Jr. et al. 2020). This backbone is operationalized by identifying the key concepts used in an interconnected way across the text. It defines interconnection as occurring when a concept appears in multiple sentences and when a given sentence includes three or more such concepts. Therefore, these interconnecting sentences bridge key concepts, and their importance can be defined in terms of the importance of the concepts they bridge, which is defined by the number of sentences in which the key concepts occur. Although this operationalization is quite straightforward, it can be challenging to implement automatically because of the complexity of natural language text.

The system uses a technology called coreference resolution to identify the same concept across the text regardless of its precise wording. For example, the first sentence in a paragraph may mention the human immunodeficiency virus but then refer to that concept later on as HIV, the virus, or it. Coreference resolution allows the proper calculation of concept importance by recognizing all these mentions as referring to the same concept. Once the important sentences have been identified, Cloze items are generated by deleting important concepts and replacing them with a blank. The important concepts are identified as previously discussed, but they are further expanded using syntactic and semantic information in the important sentences, specifically syntactic and semantic arguments.

For example, the sentence, “The unit of genetic information is a gene, which encodes a protein,” may include gene and protein as key concepts identified by coreference, but we additionally add genetic information as the object of a preposition. Thus, we would create three Cloze items for this sentence, one for each key concept. Once the questions have been generated, instructors can add, delete, or edit questions. The system also automatically generates paraphrased versions of some of the questions (Olney 2021b). These paraphrased versions are intended to reduce rote memorization from the text and improve comprehension. The automatically generated content can be selected or edited for inclusion in a class assignment.

The system provides several ways of organizing content to facilitate effective content navigation. Faculty may view automatically generated questions in the order they appear in the source text, ranked by sentence importance, coreference, or availability of paraphrased versions (Figure 1). Sorting by sentence importance allows faculty to sort the items by the AI (artificial intelligence) algorithms measure of their centrality to the text meaning. This sorting allows quick identification of the least important items, so that faculty may easily remove items that may not be necessary. Coreference sorting finds all cases where the AI algorithm detects an ambiguity in the sentence due to coreference phenomena like pronouns, e.g., “It arises from the surface of the tibia, passes _________ over the distal end of the tibia, and attaches to bones of the foot.” This item was automatically corrected to “The tibialis anterior arises from the surface of the tibia, passes _________ over the distal end of the tibia, and attaches to bones of the foot.” However, such correction is imperfect, and the sorting allows faculty to quickly identify those items that need manual checking for deletion of items or reversion of the correction. Finally, the system also paraphrases some of the sentences and offers these paraphrases as additional items.

![Figure 1](https://via.placeholder.com/150)

*Figure 1. Faculty content generation screen available to professors that shows the sorting options, editing tools, and paraphrased questions.*
Faculty can further customize their content by selecting the percentage of material made available to the students. A smaller body of questions could be more approachable for students and the percentage of material covered could be increased in updated assignments. To assign the material to the students, faculty members create a class, then assign chapters to each class. Once the material is saved and assigned to the class, it is available for students.

**Student Usage**
Currently, students log onto the system through the MoFaCTS website using their student username and password. This single sign-on authentication requires coordination with the Information Technology department at the participating school but makes it simpler for students to access the system. Once a student is logged in, they select their professor and course. Before beginning the content questions, students see directions for answering the questions and a consent form. This project was approved by the Institutional Review Board of the University of Memphis, and informed consent was obtained from all participants.

Students were asked if they had read the chapter and encouraged to do so if they had not. Rather than looking up answers, they were instructed to answer as best they could from memory. This practice was intended to move students away from hunting through the text for answers and toward recalling information from memory. Students had 30 seconds to answer a question before the question timed out and moved to the next one. Students were logged out of the system after two time-outs to prevent the system from running unattended and collecting invalid data. In the current course implementation, students were required to complete 30 minutes of practice in the chapter. Thus far, classes have required that practice be completed before the lecture exam.

Once a student begins practicing a chapter, the question sequence is controlled using an adaptive algorithm for optimal learning. This algorithm tracks student progress and uses a mathematical model to optimize when to repeat questions to improve comprehension and retention (Pavlik Jr and Anderson 2008; Pavlik Jr. and Eglington 2021a; 2021b). The algorithm uses a student’s prior performance in the mathematical model to estimate how well a student knows each Cloze item response. This estimation uses many aspects of their prior learning, including how well they have done, overall, how well they have done in prior repetitions of the Cloze, and the recency of last repetitions for each Cloze. Most importantly, it includes a representation of long-term learning as a function of the difficulty of the items since medium difficulty items (50% correct per trial) have been shown to result in the most learning (Cao et al. 2019; Pavlik Jr. et al. 2019). While the model says this is true, we also have an adjustment because errors are costlier for students (restudy takes time), which leads us to predict optimal learning at about 75% correct. During the learning session, the algorithm tries to give practice for previously seen Cloze items at this level of correctness, using the estimates for all the items to make the selection (Pavlik Jr. and Eglington 2021a). If all prior items are above this level, the algorithm can introduce new items into the currently practiced set.

If a question is answered incorrectly, the student receives feedback in one of the following forms: the correct answer, definitional feedback, refutational feedback, definitional dialogue, and refutational dialogue (Figures 2 and 3). Correct answer feedback tells the student that their answer was incorrect and provides the correct answer to fill in the blank. All other feedback types include this information but also provide additional instruction as described in the next two paragraphs.

Definitional feedback provides the student with the definitions for both the correct concept (the one they missed) and the concept they chose incorrectly. The rationale for providing information on both concepts is that the student’s error reveals three kinds of misunderstanding: the correct concept, the incorrect concept, and the relationship between them. The definitional feedback only provides information on the first two kinds of misunderstanding by drawing definitions from the textbook glossary.

Refutational feedback typically includes definitional information but focuses more on the third type of misunderstanding, the relationship between concepts. For example, if the student responded with digestive tract instead of digestive system, the refutational feedback would explain that the digestive tract is just one part of the digestive system. Refutational feedback is dynamically generated using deep learning and so has the advantage of being closely tailored to the student’s error but also the disadvantage of providing good but imperfect answers (Olney 2021a).

In addition to being available in paragraph form, both definitional and refutational feedbacks are available in tutorial dialogue form. The tutorial dialogue form is created by taking each sentence in the corresponding paragraph form and generating questions. For example, consider the questions: “What is a part of the digestive system?” or “What breaks food down into small water-soluble molecules?”. During the dialogue, 2-3 such questions are selected and presented based on the student’s ongoing answers, and feedback is given each time the student answers one of these dialogue questions. After the tutorial dialogue feedback is complete, the student returns to Cloze item practice for as long as they wish to continue.

continued on next page
Figure 2. Cloze item generated from course textbook with incorrect answer and MoFaCTS automatic answer correction.

Figure 3. Cloze item generated from course textbook that was answered incorrectly. The correct answer is cell membrane, but the student answered incorrectly with cell body (not shown). Figure shows automatically generated dialog error correction explaining the error and difference between the two structures.
Once they have completed a practice session, students can track their progress with a report generated by MoFaCTS (Figure 4). They can see the percent of items answered correctly, the number of practice items completed, and time spent in a particular section. The student report includes the “Correctness Across Repetitions of the Same Fill-in”, which is a chart showing student progress as questions are repeated for reinforcement. It also has a “Current Estimate of Recall” chart showing an estimate of recall ability for a particular question, which they can see by moving the mouse over the bar in the chart. This information gives students both numerical and graphical feedback on their progress.

### Student Survey Feedback

In continuing to develop the MoFaCTS system, student feedback is very important for us to understand how our development decisions affect students. Each semester students were offered the opportunity to take a survey about their experience with MoFaCTS. The survey asked students to rank their responses to 17 questions on a 6-point Likert scale (“Strongly Agree”, “Mostly Agree”, “Not Sure, Guess Agree” and the three corresponding disagree statements, with a final uncoded “Did not use” response). Most importantly these results can reveal insights into the student experience depending on other variables such as the course (A&P I or A&P II) in which they used the system.

According to the survey results, the perceptions of the system were marginally positive. A&P I and A&P II were analyzed separately due to differences in knowledge and education experience between two groups of students, differences that may impact how they perceived the system. Small differences in N by question were due to a handful of students who selected “did not use” for some items. It is important to note, when comparing the feedback from A&P I and A&P II students, that the A&P II students have demonstrated success initially because they were required to pass A&P I to be able to enroll in A&P II. Seven survey questions showed significant differences between A&P I and A&P II student responses (Table 1). The analysis reported below comes from a two-way ANOVA in which we controlled for the term during which the survey was submitted, since we did not want differences in the systems deployed across the two years of development to influence the comparison.

For six of these questions, students in A&P II showed significantly higher opinions regarding the system’s effectiveness. The only significant measure for which the A&P I students agreed more strongly was the statement, “I found the MoFaCTS practice items were too often about unimportant details” (Table 1). This result may reflect the fact that completing A&P I gave students more experience with studying the material and identifying important information.
for exams, resulting in their evaluating the system as more useful once in A&P II. Additionally, students who academically struggle in A&P I do not progress and would not be represented in the A&P II student population.

While this survey was not a controlled experiment, it did provide useful feedback on student experiences with the system. It is encouraging that the scores were generally positive, with A&P II students rating the system higher, as mentioned above. This indicated that the system was likely serving A&P II students better, and there is clearly a need to improve the experience for less experienced students. Since A&P I students have less prior knowledge and academic experience, they likely need the support of a tool such as MoFaCTS. This result also agrees with our overall analysis of the difficulty of practice which indicated that it was below what might be considered optimal (Pavlik Jr. and Eglington 2021a). This converging evidence that the algorithm results in practice that is too difficult has caused us to make a substantial adjustment for the final year of system testing.

**Faculty Reporting**

Faculty can view the summary tables showing class data, including the number of practice items completed, the percent correct, and the time spent by each student. Student work can be viewed by individual sections, usually chapters, and combined over the semester. Faculty can also select an individual student to see a more detailed report on their practice, either by chapter or for all content (Figure 5). The more detailed report compares an individual student’s performance with class averages in percent of questions answered correctly, practice item count, and total time spent on practice. These reports allow faculty to keep track of student progress within the system. The faculty reporting capabilities are currently being improved with feedback from previous semesters. In Spring 2022, faculty will have the ability to view completion dates for chapters, making incorporating due dates for chapter practice more feasible.

<table>
<thead>
<tr>
<th>Question</th>
<th>Term</th>
<th>AP I</th>
<th>AP II</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using the A&amp;P MoFaCTS practice exercises in my class enabled me to learn more quickly.</td>
<td></td>
<td>61</td>
<td>27</td>
<td>4.94</td>
</tr>
<tr>
<td>Using the A&amp;P MoFaCTS practice exercises enhanced my effectiveness in my class.</td>
<td></td>
<td>60</td>
<td>27</td>
<td>5.48</td>
</tr>
<tr>
<td>Using the A&amp;P MoFaCTS practice exercises made it easier to learn.</td>
<td></td>
<td>62</td>
<td>27</td>
<td>6.45</td>
</tr>
<tr>
<td>I found the A&amp;P MoFaCTS practice exercises useful in my class.</td>
<td></td>
<td>61</td>
<td>27</td>
<td>6.91</td>
</tr>
<tr>
<td>I was able to make meaningful connections to learn more deeply using the A&amp;P MoFaCTS practice exercises.</td>
<td></td>
<td>61</td>
<td>27</td>
<td>4.94</td>
</tr>
<tr>
<td>The feedback the MoFaCTS practice exercises gave was adequate to help me learn what things mean.</td>
<td></td>
<td>61</td>
<td>26</td>
<td>5.45</td>
</tr>
<tr>
<td>I found the MoFaCTS practice items were too often about unimportant details.</td>
<td></td>
<td>60</td>
<td>27</td>
<td>5.71</td>
</tr>
</tbody>
</table>

**Table 1.** End of class survey results (AP I and AP II) comparing student perceptions of the MoFaCTS system. Results given only for questions with significant differences (p<0.05). Includes sample size (N), mean (M), and standard deviation (SD).
Figure 5. Summary faculty report showing date, class, and chapter selection options, as well as percent of questions answered correctly, number of questions attempted, and time spent on practice for that chapter for each student.
Future Development

Improvements in usability for faculty and students will be made in response to continued feedback from participants. In addition to the personalization conferred by the model-based practice scheduling, we intend to further personalize practice for individual students using self-report measures collected from surveys. Survey data collected from students has shown to be predictive of student performance within the system and course exams. We found that refitting past data to include survey responses in the learner model improved accuracy. For instance, a portion of survey questions concerned the students’ self-reported reasoning for taking the course, their interest in the content, and the challenge they experienced studying anatomy and physiology. Including their answers to these questions as covariates in the learner model improved predictive accuracy.

As a concrete example, individual differences predicted different learning gains from practice and thus could inform model predictions and pedagogical decisions. Student self-reported ACT scores also predicted student performance. Students with higher self-reported ACT scores learned faster (possibly due to higher levels of prior knowledge and different study habits). Accounting for this in the model could lead to more appropriate pedagogical decisions by MoFaCTS. Students also completed an author recognition survey to estimate their reading behavior, in which they were asked to determine if presented author names were real or fake. Including their score on this test in the model also improved fit and we believe would improve the system’s efficacy. In future semesters we aim to collect this data before the students begin using the system to personalize their practice further. This approach may be especially useful early in the students’ use of the system when there is little other data yet available.

We are also continuing to improve our artificial intelligence techniques for creating Cloze items from text, paraphrases, elaborated feedback, and tutorial dialogue. Our goal is to closely match the quality of materials that A&P instructors would produce if they had the time to create such materials by hand. We have also recruited nurses to provide feedback on materials in terms of both correctness and usefulness for understanding a concept. Our ongoing efforts are focused on collecting example materials and feedback from experts that will allow us to identify weaknesses in our models so that we can be sure to provide high-quality results that are ideal for learning A&P.

Conclusion

The student survey showed that students had a positive opinion of the MoFaCTS system. A&P II students perceived the system as more valuable and more focused on important information. Over the two years of use of MoFaCTS by students, valuable information has been gained that will lead to improvements in the future. As system development continues, student perceptions of value will hopefully increase. We are planning efficacy testing in fall of 2022 to measure any direct effects of MoFaCTS on student performance. With the substantial challenges facing A&P students, effective ways to improve performance are greatly needed.

Acknowledgment

This work was supported by the Institute of Education Sciences (IES; R305A1904448). Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of IES. This work was also supported by the University of Memphis Institute for Intelligent Systems and Southwest Tennessee Community College. Special thanks to Southwest faculty members Robert Blaudow, Gayathri Kaushik, Darrick Slaughter, Marjorie Rothschild, Koushik Roy, and Sheila Bouie for their participation.

About the Authors

Amanda M. Banker, MSc, is an associate professor at Southwest Tennessee Community College, teaches anatomy and physiology, and is the A&P online course coordinator. Philip I. Pavlik Jr., PhD, is an associate professor at the Institute for Intelligent Systems and Psychology at the University of Memphis. His primary research interests include mathematical modeling of learning and memory. Andrew M. Olney, PhD, is a professor at the Institute for Intelligent Systems and Psychology at the University of Memphis. His primary research interest is natural language interfaces. Luke G. Eglington, PhD, is a postdoctoral fellow at the Institute for Intelligent Systems at the University of Memphis. His research focuses on adaptive instructional systems.

Literature Cited


continued on next page


Lauer B. 2013. Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. TESOL Quarterly 47(4):867-872. https://doi.org/10.1002/tesq.140


continued on next page
Online tutoring system (MoFaCTS) for anatomy and physiology: Implementation and initial impressions

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.68.9072&rep=rep1&type=pdf

Snow CE. 2002. Reading for understanding: Toward an R&D program in reading comprehension. Santa Monica (CA): Rand Education.

TBR: The College System of Tennessee. 2021a. General enrollment trends. Available from: https://app.powerbi.com/w?r=eyJrIjoiNTY5MTMzZWN0a2Q5eWY0YzJlNWY5ZjBkMTkyZjAtZGQ0Zi00MDQ1LTc4ZjQtZTk4Mjg0MjVhMDg5IiwidCI6Ijc4ZTkwnMDIwNDM2NjYtMzZmZS00YzFlLTQ4YzNjMTE4YTJiYmJhIiwImMiOjN9

TBR: The College System of Tennessee. 2021b. Enrollment by status and demographics. Available from: https://app.powerbi.com/w?r=eyJrIjoiMDJmMGQ2YWQyOTc1MzcyMzlkY2JlOTUwMzZiMmE5OTJjZGQ0ZS00YzFlLTc4ZjQtZTk4Mjg0MjVhMDg5IiwidCI6Ijc4ZTkwnMDIwNDM2NjYtMzZmZS00YzFlLTc4ZjQtZTk4Mjg0MjVhMDg5IiwImMiOjN9

TBR: The College System of Tennessee. 2021c. Low income. Available from: https://app.powerbi.com/w?r=eyJrIjoiNTI0YzExYTktYjI0YzExYTktYjI0YzExYTktNjQ0MzQ1NzE2MjY2IiwidCI6Ijc4ZTkwnMDIwNDM2NjYtMzZmZS00YzFlLTc4ZjQtZTk4Mjg0MjVhMDg5IiwImMiOjN9


Standing Committees:

2023 ANNUAL HOST COMMITTEE
Mark Danley
This committee is in charge of coordinating the 2023 Annual Conference to take place in Albuquerque, New Mexico

ANATOMICAL DONOR STEWARDSHIP COMMITTEE
Kelsey Stevens
This committee is charged with developing, reviewing, and recommending policies and position statements on the use of cadavers for human anatomy and physiology education in colleges, universities and related institutions.

AWARDS & SCHOLARSHIPS
Chasity O’Malley

COMMUNICATION
Larry Young
This committee is tasked with helping HAPS establish its voice in a technological landscape shaped by social media. Committee members work closely with the Marketing Committee to facilitate connections within HAPS as well as recruiting potential members via social media.

CONFERENCE
Jennifer Burgoon & Tom Lehman
This committee actively encourages HAPS members to consider hosting an Annual Conference. We provide advice and assistance to members who are considering hosting an annual conference.

CURRICULUM & INSTRUCTION
Rachel Hopp
This committee develops and catalogs resources that aid in anatomy and physiology course development and instruction.

DIVERSITY, EQUITY, AND INCLUSION
Kathy Burleson

FUNDRAISING
Stacey Dunham

MEMBERSHIP
Jacqueline Van Hoomissen

STEERING
Cindy Wingert
This committee consists of all committee chairs. It coordinates activities among committees and represents the collective committee activity to the HAPS BOD.

Click here to visit the HAPS committees webpage.

Special Committees and Programs:

HAPS EDUCATOR
Jackie Carnegie, Editor-in-Chief
Brenda del Moral, Managing Editor
This committee is responsible for publishing a quarterly edition of the HAPS Educator, the journal of the Human Anatomy and Physiology Society. The committee works closely with the Steering Committee and the President of HAPS.

EXAM PROGRAM LEADS
Valerie O’Loughlin
Dee Silverthorn
Janet Casagrand
This committee has completed, tested and approved the HAPS Comprehensive Exam for Human A&P and is developing an on-line version of the exam.

EXECUTIVE
Eric Sun
Composed of the HAPS President, President-Elect, Past President, Treasurer and Secretary

FINANCES
Ron Gerrits

NOMINATING
Kerry Hull
This committee recruits nominees for HAPS elected offices.

PRESIDENTS EMERITI ADVISORY COMMITTEE
Mark Nielsen
This committee consists of an experienced advisory group including all Past Presidents of HAPS. The committee advises and adds a sense of HAPS history to the deliberations of the BOD