Characterizing and Changing Course Elements in Undergraduate Biology Education

by

Emily Paige Driessen

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 6th, 2023

Keywords: Active Learning, Student Performance, Study Habits, Group Work, Open-Note Exams, Life Sciences

Approved by

Cissy J. Ballen, Chair, Assistant Professor, Department of Biological Sciences
Jordan Harshman, Assistant Professor, Department of Chemistry and Biochemistry
Karen McNeal, Molette Endowed Professor, Department of Geosciences
Mary Mendonca, Alumni Professor, Department of Biological Sciences

ABSTRACT

I conducted a set of four studies to catalogue and examine course elements, including active learning strategies and evidence-based teaching practices, and their influence on student experiences and performance in college biology classes. In the first study, I defined the term "active learning" and catalogued the 300+ strategies used and researched in the context of biology education research, as it was previously amorphous. Second, I conducted a meta-analysis of the effect of group work on student performance in post-secondary biology courses, showing group work has the potential to increase student performance by approximately one letter grade. Third, I investigated student perceptions of and their preparation habits for online open-note exams in an undergraduate biology class, as compared to their previous experiences with closed-note exams in other classes. Results demonstrated (1) students perceived increased exam scores, decreased exam-anxiety, decreased study time spent personally, and decreased study time spent by their peers for open-note exams, and (2) students adapted their study habits for open-note exams and students who focused on understanding, note preparation and using external resources outperformed students who did not report those study habits. Finally, I documented how the emergency transition to online classes, due to the Coronavirus disease (COVID-19) outbreak, impacted undergraduate student study habits in an introductory organismal biology class over time. We identified several consistent similarities - as well as dramatic differences - in their responses as the time away from campus increased. Together, I highlight how course elements impact student experiences and performance in post-secondary biology courses.

ACKNOWLEDGEMENTS

First, I acknowledge and thank my fantastic advisor, **Dr. Cissy Ballen**. She acted as a mentor, a psychologist, and a friend to me, and she provided me with more opportunities than I ever imagined. I thank her for accepting me into her lab, sharing her projects with me on many occasions, advancing my thinking, shaping my understanding of what and who a college biology education is for, and modeling an excellent work and life balance. I am incredibly privileged to have had the opportunity to work with such an extraordinary biology education researcher. Without her reassurance, growth mindset, constant understanding, and mentoring in both teaching and research, I would not be where I am today.

I thank my committee members **Dr. Karen McNeal**, **Dr. Jordan Harshman**, and **Dr. Mary Mendonca** for agreeing to serve on my committee and guide me through the process of earning the first biology PhD with a biology education research focus from Auburn University's Department of Biological Sciences. There were many questions involved with this process, and they helped me answer all of them.

I thank **Dr. Alan Wilson** for being an unofficial committee member. He taught me how to conduct meta-analyses and has helped me immensely with my second chapter: Group work and student performance in biology: A meta-analysis. Additionally, he supported my interest in gaining teaching experience by allowing me to co-teach his Meta-analysis class two years after I was a student in that same class. That opportunity allowed me to try new curricular materials and learn more about how best to teach a meta-analysis class.

I thank **Dr. Jenny Knight** and **Dr. Michelle Smith** for supporting me and advising me while I wrote my first biology education research publication: Chapter 1: Demystifying the

Meaning of Active Learning in Postsecondary Biology Education. I appreciated their kind words and gentle advice as I entered the field of biology education research.

A special thank you goes to my lab mates: **Dr. Abby Beatty**, **Dr. Ryan Dunk**, **Dr. Robin Costello**, **Dr. Sharday Ewell**, **Dr. Paula Adams**, **Todd Lamb**, and **Sara Odom**. Additionally, I'd like to thank the fantastic undergraduate students I had opportunity to work with: **Ian Hall**, **Peyton Brewer**, **Taylor McKibben**, **Alexis Stokes**, **Sara Wood**, **Taylor Gusler**, **Rachel Youngblood**, **and Amelia Radocha**.

I thank the department heads of biological sciences at Auburn university **Dr. Paul Cobine** and **Dr. Scott Santos**. They signed countless IRB forms on my behalf. Additionally, they helped guide me through my PhD as I figured out what classes to take as a discipline-based education researcher housed in a disciplinary biology department.

I thank my parents, **Dr. Michelle Driessen** and **Jerry Driessen**, for holding education in the highest regard. Without their encouragement at a young age, often asking me if I wanted to be a doctor, a lawyer, or a professor, I'm not sure I would have had the confidence to set my career sights this high. Additionally, they gave me one of the greatest gifts: my dog, **Lou Driessen**. Lou has been my best friend throughout my entire PhD, and he has supplied countless whines when I'm working too much and not giving him attention. He has helped keep me balanced. Pictured, I am holding a Cahaba lily and Lou.

I thank my brother, **Aaron Driessen**, for challenging me to a reading competition when I was young. I have not stopped reading and learning since. Thank you for also being a best friend in this life and a supportive and kind human

I thank **Dan Aurell** for being a constant support and ray of sunshine for me throughout this process. Thank you for cooking with and for me, listening to me when I was frustrated, and providing positive insights. With your humor and patience, you helped keep me grounded. Thank you for being a great friend to Lou and taking care of him when I had to go out of town for conferences and seminar talks.

I would also like to thank the members of the Opelika Pickleball Club. Picking up the sport of pickleball and developing a sense of community through the club did wonders for my work and life balance. I would especially like to thank **John Parks** and **Lore and Bernie Wise**.

PREFACE

In accordance with the guidelines for a PhD Dissertation in the Department of Biological Sciences at Auburn University, chapters within this dissertation are presented as stand-alone manuscripts that are either already published in or in revision with peer-review journals. Consequently, there may be some repetition in the methods of chapters and slight differences in formatting among chapters. All data chapters were conducted as collaborative research projects, primarily with my advisor, Cissy Ballen. Additional collaborators with expertise in disciplinary biology and/or biology education are included as co-authors. In all four chapters, I am the primary author and spearheaded planning and designing experiments, collecting data, analyzing data, and writing the manuscripts.

This work was conducted with the approval of the Institutional Review Board (approval numbers: Cornell IRB protocol no. 1810008360; Auburn University's IRB protocol #20–603 EX 2012; Auburn University's IRB Protocol #20-189 EX 2004.).

TABLE OF CONTENTS

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix

LIST OF TABLES

Table

LIST OF FIGURES

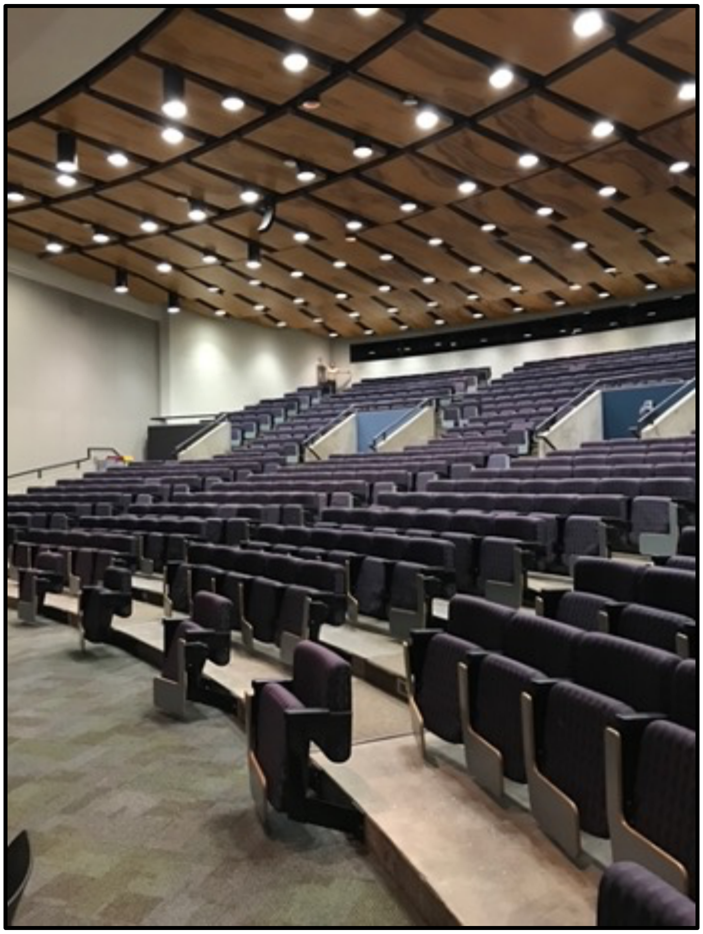
Figure

INTRODUCTION

Discipline-based education research investigates teaching and learning within STEM disciplines from a perspective that reflects the discipline's priorities, worldview, knowledge, and practices (National Research Council [NRC], 2012). The goals of discipline-based education research are to: (1) understand how people learn science and engineering concepts, practices, and ways of thinking; (2) understand the expertise of a discipline and how that expertise is developed; (3) identify and measure pertinent learning objectives and instructional approaches that lead students toward those objectives; (4) contribute to the knowledge base in a way that translates to classroom practice; and (5) make science and engineering education more broad and inclusive (NRC, 2012, p. 2). While there is much to learn and apply across the discipline-based education research fields, (e.g., biology, chemistry, physics), each subdiscipline has its own culture and is in a unique stage of development. In this dissertation, I present discipline-based education research in the field of biology: biology education research.

Biology education research is a sub-field of discipline-based education research focused on understanding biology (NRC, 2012). This subdiscipline has flourished nationally with opportunities to occupy tenure-track faculty positions, pursue PhD programs, and attend national conferences (NRC, 2012). Compared with chemistry education research and physics education research, biology education research is in a more nascent stage, only recently developing from a combination of two larger fields: (1) discipline-based education research (DBER) and (2) science education research (Dirks, 2011; Gul and Sozbilir, 2016).

A content analysis of biology education research documented trends among the research published and presented in the field. Specifically, Lo et al. (2019) analyzed the research published in *CBE – Life Sciences Education*, a key journal for biology education research



submissions, and the research presented at the Society for the Advancement of Biology Education Research (SABER) annual conference, a key annual conference for biology education research, examining the research questions, study contexts, and methods used by biology education researchers from 2002 to 2015. The analysis showed that research published in *CBE – Life Sciences Education* (*n* = 339 articles from 2002 - 2015) and presented at SABER (*n* = 652 abstracts from 2011 – 2015) focused on quantitative methods in the context of undergraduate classrooms (Lo et al., 2019).

Examples of previous biology education research focused on quantitative methods in the context of undergraduate classrooms include measuring the effect of course elements on student performance. These investigations of course elements include active learning strategies including group work (Carmichael, 2009; Chaplin, 2009; Daniel, 2016; Donovan et al., 2018; Gaudet et al., 2010; Johnson et al., 1998; Knight & Wood, 2005; Marbach-Ad et al., 2016; Springer et al., 1999; Weir et al., 2019; Yapici, 2016), clicker questions (Anderson, Schoenleber, & Korshavn, 2023; Shea et al., 2020; Smith & Knight, 2020), and flipped classrooms (Barral, Ardi-Pastores, & Simmons, 2018; Gross et al., 2015; Rai et al., 2020; Rodríguez et al., 2019). Research on course elements also examines the effect of alternative exam types such as two-stage exams/collaborative testing (Guo & Li, 2016; Meaders & Vega, 2022; Rao, Collins, & DiCarlo, 2002) or open-note exams on student performance and affect (Ambrose et al., 2010; Block 2012; Gharib, Phillips & Mathew, 2012; King, Guyette, & Piotrowski, 2009; Krasne et al., 2006; Theophilides & Koutselini, 2000; Watson & Sottile, 2010; Williams & Wong, 2007). Usually, biology education researchers conduct these experiments to meet the objectives of increasing national science literacy (Brainard, 2007; Hénard, 2009; Holm et al., 2011; Organisation for Economic Co-operation and Development [OECD], 2007; President's Council of Advisors on Science and Technology [PCAST], 2012).

The course elements investigated by biology education researchers, and by other discipline-based education researchers and education researchers, have been referred to as student-centered (American Association for the Advancement of Science, 2011), evidence-based (Moon et al., 2021; President's Council of Advisors on Science and Technology, 2012, p.iv), or active learning (Freeman et al., 2014, Theobald et al., 2020) and have been shown to increase student performance (Andrews et al., 2011; Ballen et al., 2017; Barral et al., 2018; Casper et al., 2019; Freeman et al., 2011, 2014; Haak et al., 2011; Moon et al., 2021; Theobald et al., 2020) and improve student affect (Ballen et al., 2017; Bauer et al., 2020), with some exceptions (Cooper, Downing, & Brownell, 2018). However, these investigations often (1) aren't clear about how they define active learning, (2) use small-scale investigations to evaluate active learning, student-centered, or evidence-based course elements, and (3) produce mixed-results lacking nuance. Together, these points may be used as arguments against implementing new course elements, especially for those instructors who are resistant to change (Stains et al., 2018).

In alignment with the goals and gaps of discipline-based education research, and biology education research specifically, I focus my first chapter on cataloguing definitions of active learning as well as active learning strategies used in both the biology education research literature and the biology education research community. This chapter increases clarity concerning the definition of active learning for the biology education research community. Second, I conduct a meta-analysis on the effect of group work, a commonly used active learning strategy in biology education research, on student performance in post-secondary biology classes. This chapter takes small-scale investigations and combines them to make a large-scale convincing argument that active learning strategies such as group work increase student performance in a variety of undergraduate biology classroom settings. Third, I take an evidence-based strategy, open-note exams, and investigate how performance on this type of exam is impacted by the study habits used by each student. This chapter considers the nuance of study habits used to prepare for open-note exams and finds that this nuance explains why some students may do better on these exams than others. Finally, I collect evidence on the impact of the COVID-19 pandemic on student study habits for their introductory biology exams. **The overarching goal of my research is to improve post-secondary biology education through the investigation of course elements. Specifically, my research focuses on the relationships between course elements, student performance, and student affective profiles (Figure 0.1).**

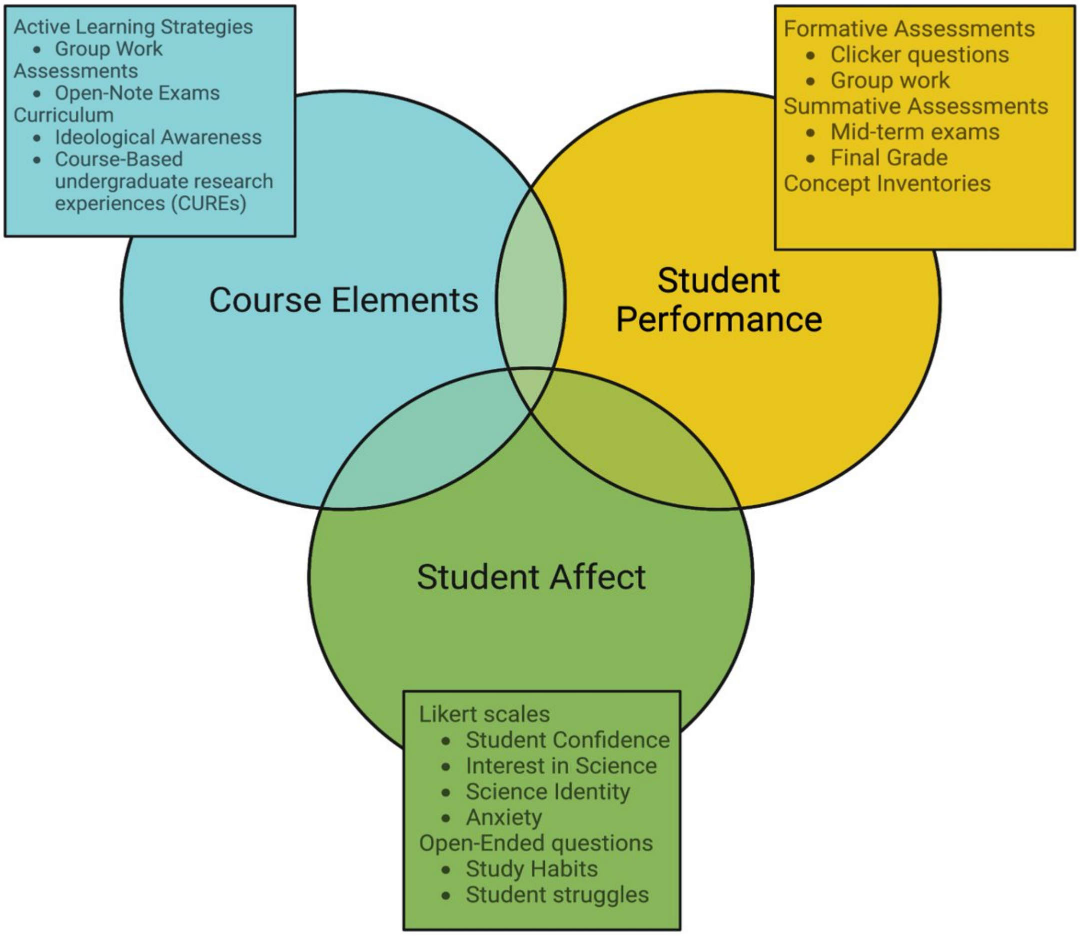


Figure 0.1 Heuristic model representing my research focus: the relationships between course elements, student performance, and student affect.

REFERENCES

Ambrose SA, Bridges MW, DiPietro M, Lovett MC, Norman MK. (2010). How learning works: Seven research-based principles for smart teaching. John Wiley & Sons.

American Association for the Advancement of Science. (2011). Vision and change in undergraduate biology education: A call to action. Retrieved October 13, 2020, from https://visionandchange.org

Anderson, D. K., Schoenleber, M., & Korshavn, S. (2023). Higher-Order Clicker Questions Engage Students and Prepare Them for Higher-Order Thinking Activities. Journal of Microbiology & Biology Education, e00151-22.

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE Life Sciences Education*, *10*(4), 394–405. https://doi.org/10.1187/cbe.11-07-0061

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education*, *16*(4), ar56.

Barral, A. M., Ardi-Pastores, V. C., & Simmons, R. E. (2018). Student learning in an accelerated introductory biology course is significantly enhanced by a flipped learning environment. *CBE Life Sciences Education*, *17*(3), 1–9. https://doi.org/10.1187/cbe.17-07-0129

Bauer, A. C., Coffield, V. M., Crater, D., Lyda, T., Segarra, V. A., Suh, K., ... & Vigueira, P. A. (2020). Fostering equitable outcomes in introductory biology courses through use of a dual domain pedagogy. CBE—Life Sciences Education, 19(1), ar4.

Block RM. (2012). A discussion of the effect of open-book and closed-book exams on student achievement in an introductory statistics course. Primus. 22(3):228–38.

Brainard, J. (2007). 50 years after Sputnik, America sees itself in another science race. Chronicle of Higher Education, 54(7), A22–A23.

Carmichael, J. (2009). Team-based learning enhances performance in introductory biology. *Journal of College Science Teaching*, *38*, 54–61. http://login.ezproxy.library.ualberta.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ofs&AN=508046050&site=ehost-live&scope=site

Casper, A. M., Eddy, S. L., & Freeman, S. (2019). True Grit: Passion and persistence make an innovative course design work. *PLoS Biology*, *17*(7), 1–10. https://doi.org/10.1371/journal.pbio.3000359

Chaplin, S. (2009). Assessment of the impact of case studies on student learning gains in an introductory biology course. *Journal of College Science Teaching*, *39*(1), 72.

Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. International Journal of STEM Education, 5(1), 1-18.

Daniel, K. L. (2016). Impacts of Active Learning on Student Outcomes in Large-Lecture Biology Courses. *The American Biology Teacher*, *78*(8), 651–655. https://doi.org/10.1525/abt.2016.78.8.651

Dirks, C. (2011). The Current Status and Future Direction of Biology Education Research. Paper presented at the Second Committee Meeting on the Status, Contributions, and Future Directions of Discipline-Based Education Research held October 18–19, 2010, in Washington, DC.

Donovan, D. A., Connell, G. L., & Grunspan, D. Z. (2018). Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. CBE Life Sciences Education, 17(4), 1–14. https://doi.org/10.1187/cbe.17-12-0283

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

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE Life Sciences Education*, *10*(2), 175–186. https://doi.org/10.1187/cbe.10-08-0105

Gaudet, A. D., Ramer, L. M., Nakonechny, J., Cragg, J. J., & Ramer, M. S. (2010). Small-group learning in an upper-level university biology class enhances academic performance and student attitudes toward group work. PLoS ONE, 5(12). https://doi.org/10.1371/journal.pone.0015821

Gharib A, Phillips W, Mathew N. Cheat Sheet or Open-Book? (2012) A Comparison of the Effects of Exam Types on Performance, Retention, and Anxiety. Online Submiss. 2(8):469–78.

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. CBE—Life Sciences Education, 14(4), ar36.

Gul, S., & Sozbilir, M. (2016). International trends in biology education research from 1997 to 2014: A content analysis of papers in selected journals. Eurasia Journal of Mathematics, Science & Technology Education, 12(6), 1631–1651.

Guo, Y., & Li, E. (2016). Collaborative Testing in Practical Laboratories: An Effective Teaching–Learning Method in Histology. Journal of Veterinary Medical Education, 43(1), 9-12.

Haak, D., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). *Increased structure and active learning reduce the achievement gap in introductory biology*. *June*, 1213–1216.

Hénard, F. (2009). Learning our lesson: Review of quality teaching in higher education draft report. Retrieved October 28, 2019, from http://www.oecd.org/education/imhe/43961761.pdf

Holm, B., Carter, V. C., & Woodin, T. (2011). Vision and change in biology undergraduate education: Vision and change from the funding front. Biochemistry and Molecular Biology Education, 39(2), 87–90.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative Learning Returns To College What Evidence Is There That It Works? *Change: The Magazine of Higher Learning*, *30*(4), 26–35. https://doi.org/10.1080/00091389809602629

King C.G., Guyette R.W. Jr, Piotrowski C. (2009). Online exams and cheating: An empirical analysis of business students' views. J Educ Online. 6(1):n1.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(4), 298–310.

Krasne S, Wimmers PF, Relan A, Drake TA. (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. Adv Heal Sci Educ. 11(2):155–71. pmid:16729243

Lo, S. M., Gardner, G. E., Reid, J., Napoleon-Fanis, V., Carroll, P., Smith, E., & Sato, B. K. (2019). Prevailing questions and methodologies in biology education research: A longitudinal analysis of research in CBE—Life Sciences Education and at the Society for the Advancement of Biology Education Research. CBE—Life Sciences Education, 18(1), ar9.

Marbach-Ad, G., Rietschel, C., & Thompson, K. V. (2016). Validation and application of the survey of teaching beliefs and practices for undergraduates (STEP-U): Identifying factors associated with valuing important workplace skills among biology students. *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-05-0164.

Meaders, C. L., & Vega, Y. (2022). Collaborative Two-Stage Exams Benefit Students in a Biology Laboratory Course. Journal of Microbiology & Biology Education, e00138-22.

Moon, S., Jackson, M. A., Doherty, J. H., & Wenderoth, M. P. (2021). Evidence-based teaching practices correlate with increased exam performance in biology. PloS one, 16(11), e0260789.

National Research Council, Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012). Discipline-based education research: Understanding and improving learning in undergraduate science and engineering (pp. 6-11). Washington, DC: National Academies Press.

Organisation for Economic Co-operation and Development. (2007). PISA 2006: Science Competencies for Tomorrow's World, Vol. 1, Analysis. Paris: OECD Publishing.

President's Council of Advisors on Science and Technology [PCAST]. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Washington, DC: U.S. Government Office of Science and Technology. Retrieved October 28, 2019, from http://www.whitehouse.gov/sites/default/files/ microsites/ostp/pcast-engage-to-excel-final2-25-12.pdf

Rai, B., Zhu, J. Y., Koh, D. C., Xiaojuan, K., Krishnaswamy, L., Chandramohanadas, R., ... & Leong, P. K. (2020). An investigation into the impact of the flipped classroom with active learning on the perception and performance of biology nonmajor students at the undergraduate level. Journal of College Science Teaching, 50(2), 27-39.

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. Advances in physiology education, 26(1), 37-41.

Rodríguez, G., Díez, J., Pérez, N., Baños, J. E., & Carrió, M. (2019). Flipped classroom: Fostering creative skills in undergraduate students of health sciences. Thinking Skills and Creativity, 33, 100575.

Shea, L., Bégin, C., Osovitz, C., & Prevost, L. (2020). Comparison of student success between high-clicker and low-clicker frequency in a large-enrollment introductory biology course. Journal of College Science Teaching, 50(2), 78-87.

Smith, M. K., & Knight, J. K. (2020). Clickers in the biology classroom: strategies for writing and effectively implementing clicker questions that maximize student learning. Active learning in college science: the case for evidence-based practice, 141-158.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21–51. https://doi.org/10.3102/00346543069001021

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... & Young, A. M. (2018). Anatomy of STEM teaching in North American universities. Science, 359(6383), 1468-1470.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proceedings of the National Academy of Sciences, 117(12), 6476-6483.

Theophilides C, Koutselini M. Study behavior in the closed-book and the open-book examination: A comparative analysis. Educ Res Eval. 2000;6(4):379–93.

Watson GR, Sottile J. (2010). Cheating in the digital age: Do students cheat more in online courses?

Weir, L. K., Barker, M. K., McDonnell, L. M., Schimpf, N. G., Rodela, T. M., & Schulte, P. M. (2019). Small changes, big gains: A curriculum-wide study of teaching practices and student learning in undergraduate biology. *PLoS ONE*, *14*(8), 1–16. https://doi.org/10.1371/journal.pone.0220900

Williams JB, Wong A. (2007). Closed book, invigilated exams versus open book, open web exams: An empirical analysis. ICT Provid Choices Learn Learn Proc Australas Soc Comput Learn Tert Educ Singapore. 1079–83.

Yapici, İ. Ü. (2016). Effectiveness of Blended Cooperative Learning Environment in Biology Teaching: Classroom Community Sense, Academic Achievement and Satisfaction. *Journal of Education and Training Studies*, *4*(4), 269–280. https://doi.org/10.11114/jets.v4i4.1372

SUMMARY OF CHAPTERS

Chapter 1: Demystifying the Meaning of Active Learning in Postsecondary Biology Education

Active learning is frequently used to describe teaching practices, but the term is not well-defined in the context of undergraduate biology education. To clarify this term, we explored how active learning is defined in the biology education literature (*n* = 148 articles) and community by surveying a national sample of biology education researchers and instructors (*n* = 105 individuals). Our objectives were to increase transparency and reproducibility of teaching practices and research findings in biology education. Findings showed the majority of the literature concerning active learning never defined the term, but the authors often provided examples of specific active-learning strategies. We categorized the available active-learning definitions and strategies obtained from the articles and survey responses to highlight central themes. Based on data from the BER literature and community, we provide a working definition of active learning and an Active-Learning Strategy Guide that defines 300+ active-learning strategies. These tools can help the community define, elaborate, and provide specificity when using the term active learning to characterize teaching practices.

Chapter 2: Group work and student performance in biology: A meta-analysis.

Active learning is hailed as the pinnacle of teaching in STEM classrooms for its ability to maximize learning. While previous work broadly defined active learning as the antithesis of lecture, measuring the effects of all teaching strategies combined, we tested the isolated impacts of group work on student performance, using meta-analysis. We collected data from 34 articles representing 66 estimates calculated from performance outcomes of 18,494 students. Our overall estimate indicates the implementation of group work in biology classrooms increased student performance by 0.72 standard deviations. When applied to a grade scale based on real-world data, the magnitude of this change is roughly one letter grade. Moderator analyses revealed the increase in performance holds across all class sizes, class levels, and group sizes up to five students. These results demonstrate group work leads to impressive boosts in student performance and reveal the value of studying specific active learning strategies.

Chapter 3: Evaluating open-note exams: student perceptions and preparation methods in an undergraduate biology class

Although closed-note exams have traditionally been used to evaluate students in undergraduate biology classes, open-note exams are becoming increasingly common, though little is known about how students prepare for these types of exams. We investigated student perceptions of and their preparation habits for online open-note exams in an undergraduate biology class, as compared to their previous experiences with closed-note exams in other classes. Specifically, we explored the following research questions: (1a) How do students perceive open-note exams impact their exam scores, their anxiety, the amount they studied, and the amount their peers studied? (1b) How do these perceptions impact performance outcomes? (2a) How do students prepare for open-note exams? (2b) How do these preparation methods impact performance outcomes? Results demonstrate students perceived increased exam scores, decreased exam-anxiety, decreased study time spent personally, and decreased study time spent by their peers for open-note exams, as compared to past experiences with closed-note exams. Open-ended survey responses analyzed through first- and second-cycle analyses showed students adapted their study habits by focusing on note preparation and broad conceptual understanding rather than rote memorization. Using linear mixed effects models to assess student performance, we found students who focused on understanding, note preparation and using external resources outperformed students who did not report those study habits. As institutions shift towards flexible and scalable assessments that can be used in face-to-face or online environments, the use of open-note exams can promote effective study habits and reward higher-order thinking with intentional guidance from the instructor.

Chapter 4: Learning principles of evolution during a crisis: An exploratory analysis of student barriers one week and one month into the COVID-19 pandemic

The Coronavirus disease (COVID-19) outbreak forced an emergency transition to online classes across the world with little warning or instruction for faculty and students. The goal of this research was to document how this response impacted undergraduate students studying the principles of evolution in an introductory organismal biology class over time; specifically, how their study habits for exams differed (1) one week and (2) one month after a university's decision to transition to emergency remote instruction. We asked students about the extent to which COVID-19 impacted their study habits, and we categorized students' responses using open coding. We identified a number of consistent similarities - as well as dramatic differences - in their responses as the time away from campus increased. The report that follows is a summary of the documented barriers and recommendations based on literature concerning crises and equitable practices.

Chapter 1: Demystifying the Meaning of Active Learning in Postsecondary Biology Education

Emily P. Driessen[1], Jennifer K. Knight[2], Michelle K. Smith[3], Cissy J. Ballen[1]

1. Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA; 2. Department of Molecular Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA; 3. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

Corresponding author: Emily Driessen (epd0016@auburn.edu)

Keywords: biology, active learning, strategies, definitions

Manuscript published as a Research Article in CBE—Life Sciences Education in 2020.

ABSTRACT

Active learning is frequently used to describe teaching practices, but the term is not well-defined in the context of undergraduate biology education. To clarify this term, we explored how active learning is defined in the biology education literature (n=148 articles) and community by surveying a national sample of biology education researchers and instructors (n=105 individuals). Our objectives were to increase transparency and reproducibility of teaching practices and research findings in biology education. Findings showed the majority of the literature concerning active learning never defined the term, but the authors often provided examples of specific active learning strategies. We categorized the available active learning definitions and strategies obtained from the articles and survey responses to highlight central themes. Based on data from the biology education research literature and community, we provide a working definition of active learning and an Active Learning Strategy Guide that defines 300+ active learning strategies. These tools can help the community define, elaborate, and provide specificity when using the term active learning to characterize teaching practices.

BACKGROUND

The promotion of undergraduate biology knowledge in the United States has immediate and long-term implications for increasing national science literacy, providing high-quality education to the STEM workforce, and contributing to critical scientific advances. To meet these objectives, calls to action formalized priorities and made specific recommendations aimed at improving undergraduate biology education nationwide. For example, after extensive discussions among biology faculty, students, and administrators, the American Association for the Advancement of Science (2009) published a formative document, *Vision and Change: A Call to Action*, which advocated for “student-centered classrooms” and outlined six core competencies intended to guide undergraduate biology education: (1) apply the process of science; (2) use quantitative reasoning; (3) use modeling and simulation; (4) tap into the interdisciplinary nature of science; (5) communicate and collaborate with other disciplines; and (6) understand the relationship between science and society. Another call to action came from the President’s Council of Advisors on Science and Technology, who proposed five recommendations to change undergraduate STEM education, including the adoption of “evidence-based teaching practices” (Olson & Riordan, 2012).

Although these pushes for “student-centered” and “evidence-based” practices are relatively recent, they stem from ideologies that are more than a century old. Specifically, Dewy (1916) wrote, “learning means something which the individual does when he studies. It is an *active*, personally conducted affair” (p. 390). Based upon this work, Pesavento, et al. (2015) identified Dewy as one of the earliest and most influential advocates of what we now know as active learning. Subsequently, others expanded on and institutionalized terms such as “student-centered” and “evidence-based” practices (Piaget, 1932; Montessori, 1946; Vygotsky, 1987;

Papert, 1980; Brown, Collins & Duguid, 1989; Turkle & Papert, 1990; Ackermann, 2001, Cook, Smith, & Tankersley, 2012). While this body of work is critical to our understanding of active learning, the ways in which practitioners and researchers currently use the term are often vague.

Despite this ambiguity, research concerning the effectiveness of active learning in the classroom has continued. For example, a landmark meta-analysis compared student achievement and failure rates between undergraduate science, engineering, and mathematics classes that used active learning approaches and those that used lecture (Freeman et al., 2014). Findings demonstrated that active learning decreased failure rates by 55% and increased student examination performance by approximately half a standard deviation. To define active learning for the purposes of clarity and transparency in their research, Freeman et al. (2014) developed a definition based on responses from 338 biology departmental seminar audience members: "Active learning engages students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert. It emphasizes higher-order thinking and often involves group work" (pgs. 8413-8414). This definition guided their inclusion criteria for the study, and it is one of the few examples of clearly defined parameters.

Although many articles do not define the exact parameters of active learning, the research has demonstrated the positive effects of active learning on student achievement and affect across multiple contexts. For example, researchers demonstrated that active learning yields disproportionate learning gains among the most at-risk student groups such as first-generation college attendees and those who identify with race/ethnicities historically underrepresented in STEM fields (Ballen, Wieman, Salehi, Searle, & Zamudio, 2017; Bauer et al., 2020; Beichner et al., 2007; Eddy & Hogan, 2014; Haak, HilleRisLambers, Pitre, & Freeman, 2011; Wilton et al., 2019). Additionally, a meta-analysis conducted by Theobald et al. (2020) demonstrated that active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and mathematics disciplines. However, it is important to note that the definitions of active learning used in these articles vary from the antithesis of lecture (Theobald et al., 2020) to listing the specific strategies that characterize the term (e.g. in-class activities, pre-lecture preparation, and frequent low-risk assessment; Ballen et al, 2017).

Despite the varying parameters of the term, post-secondary institutions have increasingly embraced the use of the term active learning (Pfund et al., 2009; Aragón et al., 2018). Examples include institution-wide initiatives (e.g., the Science Education Initiatives at University of Colorado and University of British Colombia, and the Active Learning Initiative at Cornell University), the Summer Institutes on Scientific Teaching (https://www.summerinstitutes.org), and the Obama Administration's Active Learning Day (https://obamawhitehouse.archives.gov/blog/2016/10/25/active-learning-day-america). Additionally, more than three-fourths of colleges and universities in the United States provide some type of active learning classrooms, defined as those that offer flexibility in design to facilitate different types of teaching (Alexander et al., 2019).

Despite these institutional supports and documented positive impacts, the term active learning itself is difficult to ascertain from a review of literature. For example, Eddy, Converse, and Wenderoth (2015) explained active learning is a complex process that encompasses both teaching methods and student learning. Drew and Mackie (2011) noted the meaning of active learning may be dichotomous as it has been considered a theory of learning as well as a set of pedagogical strategies. Although attempts have been made to define active learning as a theory (Freeman et al., 2014; Connell, Donovan & Chambers, 2016; Moss-Racusin et al., 2016; Auerbach & Schussler, 2017; Jeno et al., 2017) as well as a set of strategies in biology education research (BER) (Tanner, 2013; Miller & Tanner, 2015), these attempts are not always (1) streamlined or easy to follow, (2) regularly used in the literature, (3) supported by literature or data, and/or (4) comprehensive. This outcome is problematic when trying to understand what exactly active learning encompasses.

Notably, the variation in the conceptualization of active learning reflects a state of scientific revolution. According to Kuhn (1970), the development of a science has alternating phases (i.e. normal and revolutionary). Normal science, equated to puzzle-solving, comes with a reasonable chance of solution via familiar methods and can be solved by one person. On the other hand, a revolutionary phase involves a collectively negotiated revision to an existing belief or practice. While discipline-based education researchers address questions about the efficacy of recently-developed teaching strategies, those strategies are commonly being binned under active learning, which is an ill-defined term. In order to improve our field, it is important to negotiate how the community interprets and understands this term.

Furthermore, demystifying active learning in undergraduate biology has direct applications for teaching and research. The broad interpretation of active learning may discourage instructors from trying new instructional practices and ultimately serve as a barrier to implementation (Kreber & Cranton, 2000; O'Donnell, 2008; Stains & Vickrey, 2017). It may additionally serve as a barrier to experimental replication in discipline-based education research communities because there are not agreed upon standards, or criteria for inclusion or exclusion. Given this, we investigated the following four questions in the context of undergraduate biology courses: (1) How does the BER literature use and define the term active learning? (2) How does the BER community define the term active learning? (3) How are active learning strategies described in the BER literature? and (4) How are active learning strategies described by the BER community? We addressed these research questions through a review of BER literature and a survey of the BER community. We expect that by developing ways to efficiently communicate active learning in the context of biology education, we will encourage teaching innovations and the adoption of common research-based practices.

METHODS

Analyzing the Literature

To address how the BER literature defines and uses the term active learning, we extracted information from peer-reviewed biology education journals. Many peer-reviewed journals publish biology education research, including *Advances in Physiology Education, The American Biology Teacher, Anatomical Science Education, BioScience, Journal of College Science Teaching*, *Life Sciences Education,* and the *Journal of Microbiology & Biology Education*; however, we chose to examine only two of them, acknowledging that this is an exploratory, non-exhaustive study. We chose *Life Sciences Education* and the *Journal of Microbiology & Biology Education* because of their prominence, history, and readership (see websites; https://www.lifescied.org/; https://www.asmscience.org/content/journal/jmbe). We searched for the term 'active learning' in the titles, abstracts, or text of research articles published in those two journals, and only papers that used this term were included in our analysis. In order to get a contemporary snapshot of how the term is used, we only included articles published over the three years that preceded the start of the study, from January 1, 2016 to December 31, 2018. We collected data within the same time span from *CourseSource,* an online journal that exclusively publishes evidence-based biology teaching materials for undergraduate classrooms and laboratories (https://www.coursesource.org/about). We included this journal because it captures how biology instructors translate findings from the active-learning research literature into classroom practice. All *CourseSource* lesson articles included an Active Learning section where authors list and/or explain their instructional approaches, so we included all published papers in the final analysis.

Once we selected articles based upon our search criteria, three of the authors (C.J.B., M.K.S., J.K.K.) read the articles and extracted the relevant text surrounding the search term active learning. If active learning was defined in the manuscript, it usually occurred after the term was first mentioned in the introduction or the methods. Articles that described specific active learning strategies often included them in the methods section of the paper, after introducing active learning broadly as an effective form of instruction. Because this text placement could vary, we searched through each manuscript to make sure we included any definition or strategies that the manuscript's authors described.

To determine how biology education research articles use and define the term active learning, we first examined to what extent, if at all, articles included a definition of active learning. Articles that met the inclusion criteria were binned into six categories first based on whether researchers followed their definition of active learning with a citation (i.e., they were 'literature-based'), or if researchers did not include a citation (i.e., they were 'not literature-based'). Second, we recorded whether articles included specific active learning strategies, either in addition to a formal definition or in place of a definition. Our final list of categories included articles that provided: (1) a definition of active learning that was literature-based (i.e., included a citation) with examples of active learning strategies; (2) a definition of active learning that was *not* literature-based (i.e., did not include a citation) with examples of active learning strategies; (3) a definition of active learning that was literature-based with no examples of active learning strategies; (4) a definition of active learning that was *not* literature-based and had no examples of active learning strategies; (5) no active literature definition with active learning strategies; and (6) no active literature definition and no active learning strategies (Figure 1.1).

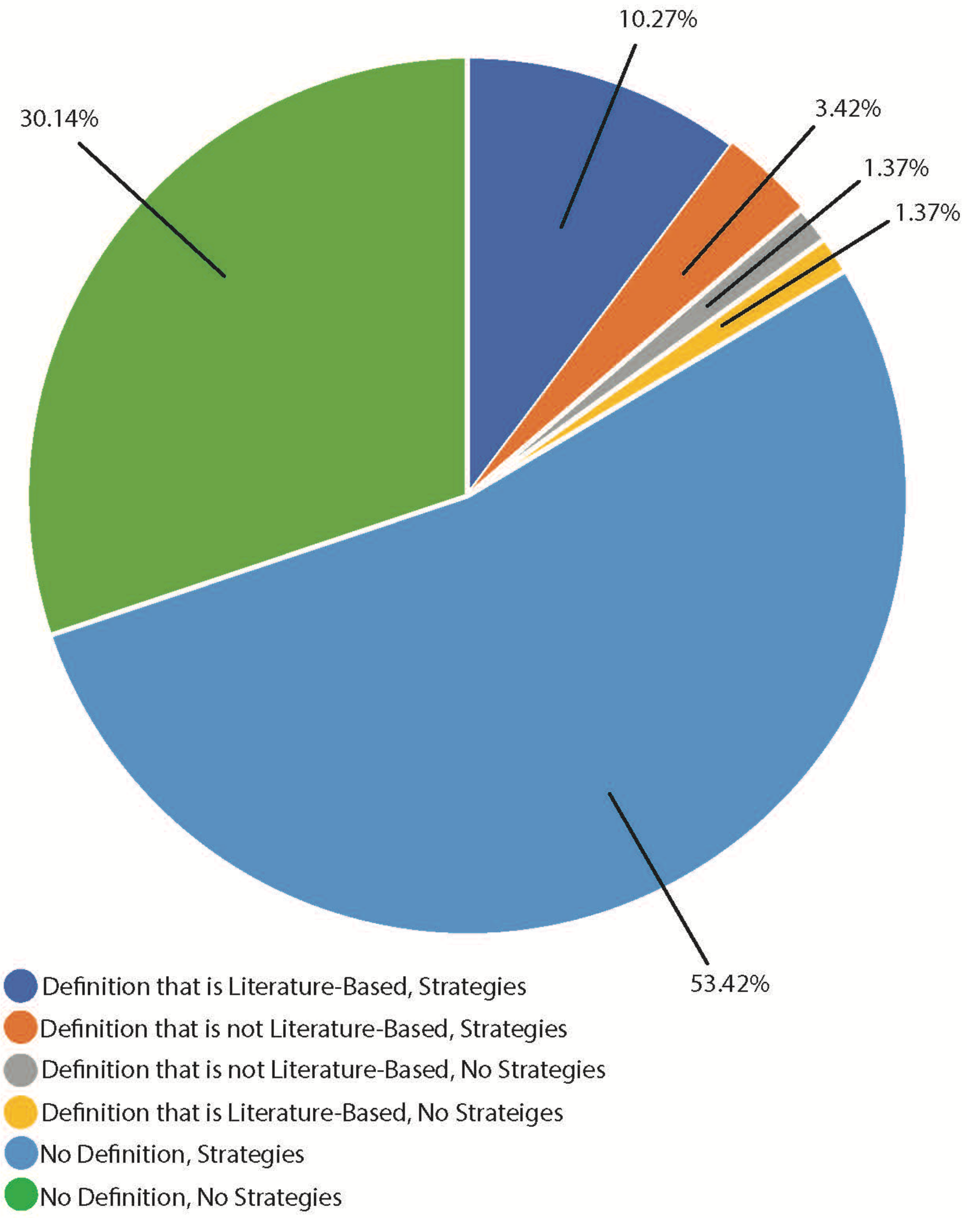


FIGURE 1.1 Ways in which articles from *LSE*, the *Journal of Microbiology & Biology Education*, and *CourseSource* use the term “active learning.”

Surveying the Community

In addition to combing the literature, we collected survey data from members of the Society for the Advancement of Biology Education Research (SABER; a scientific community of discipline-based education researchers and teaching practitioners who focus on improving post-secondary biology education through evidence and theory) via the listserv. We selected SABER as a group to survey because it is the "world's largest organization dedicated to scientifically exploring how to teach biology most effectively" (website: https://saberbio.wildapricot.org/).

Through the survey, we collected demographic information from the survey participants, including their institution type, employment position (i.e., faculty, postdoc, graduate student, etc.), level of biology class (e.g. lower level, upper level, etc.), class size, country of instruction, and frequency of active learning instruction practice (Table 1.1). Additionally, the survey included the following two prompts: (1) in your own words, define the term active learning in the context of undergraduate biology classrooms; and (2) list the active learning techniques that you use in biology classrooms. All research was conducted in accordance with the Cornell University Institutional Review Board (Cornell IRB protocol # 1810008360).

TABLE 1.1 Information about SABER members who participated in the survey

How would you describe your institution type?
PhD-granting institution (60%)
Primarily undergraduate institution (17%)
Community college (14%)
Master's-granting institution (5%)
Other (4%)

What is your current position?
Faculty (71%)
Other (14%)
Postdoc (11%)
Graduate student (4%)
What type of biology class do you teach? (Select all that apply)
Combination of lower level, upper level, or graduate (58%)
Lower level (introductory; 26%)
Upper level (requires at least one prerequisite; 14%)
Graduate level (2%)
What size biology classes do you teach? (Select all that apply)
Small (50 or fewer students; 42%)
Combination of small, medium, and large (38%)
Large (more than 100 students; 14%)
Medium (more than 50 up to 100; 6%)
Do you use active-learning instruction practices in your teaching?
Yes, in every class period (86%)
Yes, in some class periods (14%)
No (0%)

Data Categorization

After we obtained both active learning definitions and strategies from the surveys and the literature, we analyzed the data. Specifically, we started by creating two data sets. These were created by (1) taking the active learning definition text from the literature and from the surveys and combining it into one Excel spreadsheet, and (2) taking the active learning strategies text from the literature and from the surveys and combining it into another Excel spreadsheet. Both data sets were then categorized.

Categorization Active Learning Definitions

After combining the active learning definitions from both data sources (i.e., the literature and surveys), we analyzed the active learning definition data. Using open coding (Strauss, 1987; Strauss and Corbin, 1990), a method rooted in the grounded theory framework (Glaser, Strauss & Strutzel, 1968), three of the authors (J.K.K., M.K.S. & C.J.B.) reviewed the responses and identified recurring themes. Using the methodology from Saldaña (2015), the authors compared their notes and developed a final set of ten categories: students interacting or engaging with the material, not traditional lecture, group work, scaffolding or constructivism, problem-solving, individual formative assessments (e.g., through the use of personal response systems), student-centered pedagogy, application or synthesis of material, student ownership of learning, and evidence-based teaching. The authors binned each article's use of the term into as many categories of active learning as appropriate. At first, the three coders placed 71% of the definitions in the same categories. After discussion, coders resolved all differences and shared 100% agreement. Then, we calculated how often the definitions appeared in each of the ten categories, for the surveys and the literature respectively.

Categorization Active Learning Strategies

After merging the strategies from the literature with those obtained from the surveys, we analyzed the active learning strategy data. Using open coding, three researchers (E.P.D. and two undergraduate students) developed a set of nine categories (Appendix A): metacognition, discussion, group work, assessment, practicing core competencies, visuals, conceptual class design, paperwork, and games. To improve our collective ability to reliably categorize strategies, we needed definitions of each strategy listed. Because no such list of definitions existed, we defined each of the unique strategies (Appendix B) using published literature or dictionary definitions (Appendix C). The utility of this list, which we call the Active Learning Strategy Guide, can also be used by the education research community and disciplinary practitioners interested in learning about active learning strategies. Using the Active Learning Strategy Guide, we were able to categorize the strategies with an initial percentage of agreement of 75%. After discussion, the researchers resolved any differences with discussion for a final percentage of agreement of 100%.

RESULTS

How does the BER literature use and define the term active learning?

Of the 148 articles that fit our search criteria, the majority did not provide a definition for the term active learning; but instead listed examples of specific active learning strategies (53.42%; Figure 1.1). The second most common approach used in the articles provided less information; no definition and no list of relevant strategies (30.14%). Overall, this demonstrates the overwhelming majority of the active learning literature (83.56%) did not define active learning.

To address how the biology education research literature defines active learning, we focused on articles that provided a definition of the term, with or without the inclusion of one or more references. Among the 24 articles that defined active learning (Table 1.2), 17 articles (74%) provided literature citations and seven (26%) did not. Of the 17 articles that defined active learning using references to the literature, 5 of them (29%) cited Freeman et al. (2014). There was a bit of variation in reference use though, with a total of 43 different references mentioned (Appendix D).

Table 1.2 Active-learning definitions pulled from the literature. Citations available in Supplemental Appendix D.

Citation	Active Learning Definition
Auerbach & Schussler (2017)	Active learning is a *student-centered* pedagogical approach that engages student thinking through the use of class activities that require students to reflect upon and often explicitly discuss their ideas and their application. [1, 2]
Ballen et al. (2017)	…characterized by in-class activities, pre-lecture preparation, and frequent low-risk assessment…[3, 4, 5, 6, 7]
Becker et al. (2017)	…active-learning…requires students to take responsibility for their own knowledge level. [8]
Bentley & Connaughton (2017)	Active learning occurs whenever students participate in an activity that allows them to process or synthesize course content.
Bouwma-Gearhart et al. (2018)	Part of student engagement (i.e. active learning is anything that engages students).
Cavanagh et al. (2018).	Active learning involves a range of student-centered pedagogies that encourage engagement through activities such as peer collaboration, experimentation, writing, and problem solving. [9, 10]
Connell, Donovan & Chambers (2016)	Active-learning pedagogies are intended to move classrooms toward more student-centered learning, and they engage students in knowledge construction. This context is in contrast to traditional lecture, which focuses on dissemination of instructor knowledge and relies on passive student listening.
Cooper, Ashley & Brownell (2017)	In contrast to instructors predominantly transmitting information to students by lecturing during class, active-learning classrooms give students the opportunity to construct their own knowledge, often through group work with other students. [7, 8, 11, 12, 18]
Cooper & Brownell (2016)	…in active-learning classes, students are asked and often required to actively engage with other students and the instructor…there is not a single, agreed-upon definition of active learning.[7]
Cooper et al. (2018)	…students engage in constructing their own knowledge during class. [7, 13, 14, 15]
Cooper et al. (2017)	In contrast to traditional lecture, students in active-learning classrooms are expected to interact more frequently with one another as well as with the instructor.[12, 16, 17, 18]
Dewsbury (2017)	Active learning is a broad concept.
Durham, Knight & Couch (2017)	…students are explicitly asked to engage in thinking about course material during class.[19, 20, 21]
Elliott et al. (2016)	Active-learning strategies require students to engage with concepts and then provide students with feedback on their learning process.
Goff et al. (2017)	…the active engagement of students in the learning process…[13, 23]
Green, McMahon & Brame (2016).	...the main aspects of active learning: learning by doing and metacognition.
Hoefnagels & Taylor (2016)	Active learning is difficult to define, but the overall goal is simple: to reduce the amount of time that students spend passively listening to lectures.
Jeno et al. (2017)	…active learning requires students to actively interact with the learning material and has been shown to have a positive effect on retention, as well as reducing dropout and failure rates. [7, 24]
Kudish et al. (2016)	…require critical integrative and analytical thinking. [7, 25, 26]
Lee et al. (2018)	Active learning…focuses on moving students from passively receiving and replicating instruction to actively using, testing, and formulating what they are learning for themselves....
McCourt et al. (2017)	…active learning…can be defined as pedagogies that require students to engage directly in and take responsibility for their own learning. In practice, active learning occurs when

	instructors stop lecturing and provide time for students to complete activities that build conceptual understanding.
McLean & Suchman (2016)	Active learning advocates contend that when students do something they learn it better than if they just hear and see it.[27]
Moss-Racusin et al. (2016)	…a collection of teaching methods that *engage* learners and provide practice in scientific thinking.[9, 28, 29, 30, 31, 32]
Stoltzfus & Libarkin (2016)	Active learning is based on *constructivist* theory—the idea that students must create their own knowledge in order for learning to persist. One core feature of active learning in the classroom is a *decrease in lecturing* during which students passively listen and an increase in outcome-related activities in which students actively develop their own understanding.[1, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43]

The seventeen definitions obtained from the literature were categorized as previously mentioned. The most represented category defined active learning as "students interacting or engaging with the material," followed by the category that emphasizes what active learning is *not*: "not traditional lecture" (Figure 1.2A).

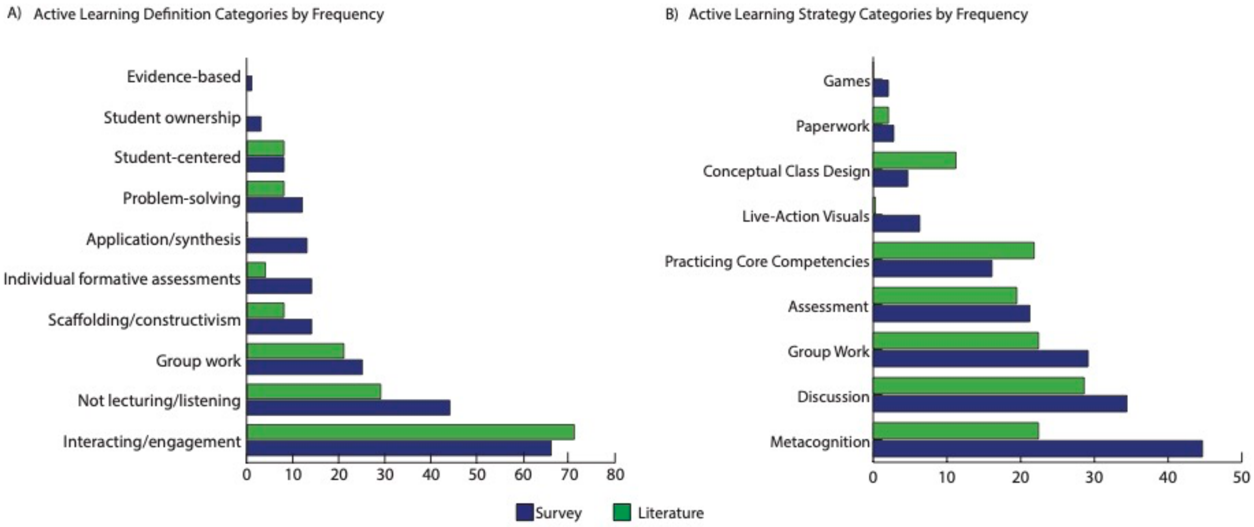


FIGURE 1.2 Frequency of how the BER literature and community define active learning an describe strategies. (A) The categorized definitions of active learning from the literature (*LSE*, *Journal of Microbiology & Biology Education,* and *CourseSource*) and a survey disseminated to SABER members. Each bar shows the percentage of articles or total survey respondents who included the corresponding term in their definition of active learning or list of active-learning strategies. (B) The categorized active-learning strategies from the same BER

literature and community sources. The graph is organized by increasing percentage of total survey responses in each category. The percent values represented in each figure do not add up to 100%, because each literature source and survey response could have more than one strategy or definition represented.

How is the term active learning defined by the BER community?

We received responses from 105 individuals from a variety of institutions across the United States (Table 1.1). In general, survey participants' definitions fit into the same categories as those in the surveyed published literature (Figure 1.2A). The most common definition of active learning, from the BER community, was "interacting/engagement" with the material. The second most common categorized definition was "not lecturing/listening," followed by "group work."

How are active learning strategies described in the BER literature?

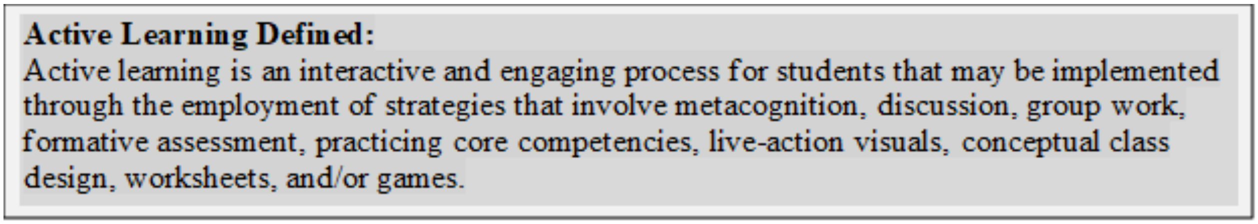
After analyzing the qualifying articles, we found that 38% of them did not mention any specific active learning strategy. From the papers that mentioned active learning strategies, a total of 339 strategies were extracted with 133 of them being unique responses. Once these strategies were categorized, the data revealed the most frequently represented strategy categories from the literature were discussion (29%), group work (22%), and metacognition (22%; Figure 1.2B).

How are active learning strategies described by the BER community?

We asked survey participants to respond to the prompt, "List the active learning strategies that you use in biology classrooms." We collated a list of 681 strategies from the responses, of which 201 were unique. After categorizing these strategies, we found the most frequently represented strategy categories from the surveys were discussion (34%), group work (29%), and metacognition (45%) (Figure 1.2B).

DISCUSSION

Our aim was to bring clarity and transparency to the term active learning as it is used within the biology education research community. We addressed this by identifying the definitions and strategies attributed to the term by analyzing the literature and surveying a biology education research society. From these compiled findings, we constructed an active learning definition (see box "Active Learning Defined") as well as a reference guide for 300+ defined active learning strategies (Appendix B).



Active Learning Defined:
Active learning is an interactive and engaging process for students that may be implemented through the employment of strategies that involve metacognition, discussion, group work, formative assessment, practicing core competencies, live-action visuals, conceptual class design, worksheets, and/or games.

Below we propose future steps for the biology education research community with accompanying tools to aid in the process. First, we advocate for all biology education research concerning active learning to provide a cited definition. Second, we suggest authors define and describe the active learning strategies utilized in the experimental research.

Define the Term

Active learning has rarely been defined in the literature. This outcome could be due to the lack of a unanimous definition for the biology education research community, the fact that active learning is a complex process encompassing both teaching methods and student learning (Eddy et al. 2015), the dichotomous nature of the term as both a theory and as a set of pedagogical strategies (Drew and Mackie, 2011), the perception that this term is self-descriptive, and/or the notion that it is unimportant given the majority of the research articles focused on the effects of the implementation of a specific active learning strategy. Whatever the reason, we advocate for the inclusion of definitions in BER articles in order to clarify the author's interpretation. This is because, based on our investigation into the contemporary literature, it is apparent that people interpret active learning in a variety of different ways (i.e., interacting/engagement, not lecturing/listening, group work, scaffolding/constructivism, individual formative assessments, application/synthesis, problem-solving, student-centered, and evidence-based). Ultimately, providing a definition may aid in increased fidelity and reproducibility of experimental outcomes (Stains & Vickrey, 2017).

When considering outcomes such as those demonstrated by Freeman et al (2014) (i.e. active learning decreased failure rates by 55% and increased student examination *performance by approximatel*y *ha*lf a standard deviation), increased fidelity and reproducibility of experimental outcomes is important, especially since the promotion of undergraduate biology knowledge in the United States is consequential to critical scientific advances. To help in these efforts, we provide a number of resources and suggestions. First, we provided all of the definitions of active learning collected from recent biology education research literature in addition to the references used to support them, when applicable (Table 1.2). We also constructed a working definition of active learning based on the summarized input from the 148 articles found in the biology education research literature and the 105 responses from the BER community. This definition can be used confidently by the biology education research community in their own research given it is based on an average of biology education research literature and instructor responses.

Define the Strategies

The research papers we examined commonly listed active learning strategies. Many of the strategies were either (1) self-descriptive; the meaning could be easily deciphered from the term (e.g. applying knowledge of other subjects, circulate to check for understanding, group brainstorming); (2) defined in the literature by Tanner (2013), Miller and Tanner (2015), or others; or (3) easily collapsed into one of the three most common categories (i.e., metacognition, group work, or discussion). However, many strategies lacked transparency because authors did not describe how they implemented them. We found these cases problematic because the strategies would be difficult to replicate. To improve clarity and transparency, we share with readers our comprehensive list of unique strategies, collected from both the literature and the surveys, with definitions from the literature when available (Appendix B), as well as citations of articles in which they were used in practice (Appendix D).

Additionally, we have created a living document version of Appendix B that can be viewed using the following link: https://www.ballenlab.org/active-learning-strategies-in-biolo. Contributions or constructive feedback from the community is welcome; you can make a submission by contacting the lead author or using the following Google form: https://forms.gle/Boh6NNm1rqzHACXi8. This feedback will be considered and used by the lead author to improve the living document going forward. Our hope is that biology education researchers and teachers use these tools to define active learning strategies they have used or as guides to articles that previously implemented these strategies. It is important to note that the strategies used and the efficacy measured in those studies may vary based on fidelity of implementation.

Another way to increase the precision of descriptions is the use of observation protocols that can characterize classroom instruction behaviors. Some examples include the Teaching Dimensions Observation Protocol (Hora et al. 2013), the Classroom Observation Protocol for Undergraduate STEM (Smith et al. 2013), the Practical Observation Rubric To Assess Active Learning (Eddy et al. 2015), and the Measurement Instrument for Scientific Teaching (Durham et al. 2017). These protocols document the frequency of multiple instructional practices, include categories of active learning strategies, and can be helpful both for research purposes and to provide feedback to instructors on their practices. Such information can provide valuable guidance to biology educators, especially when used in conjunction with data on student performance, attitudes, social psychological factors, and self-reflective practices.

Limitations and Future Work

One limitation of this work is that the active learning definitions and strategies were solicited from the biology education research community only. While we hypothesize that these definitions and strategies may overlap with other discipline-based education research subjects (e.g. chemistry, geology, physics, etc.), we cannot generalize our results across disciplines, given results from Lund and Stains (2015) revealed differences in the factors influencing the adoption of evidence-based instructional practices among disciplinary chemistry, biology, and physics faculty. However, many of the strategies featured in the Active Learning Strategy Guide may be useful across disciplines. Additionally, it is reasonable to expect we may have received different active learning definitions and strategies from disciplinary biology instructors or teaching practitioners whom do not have a biology education research background. While seeking that information is out of the scope of this research, BER would benefit from engaging with the larger community to see how their work is translated among practitioners.

Second, while it is important to understand how active learning is used in classroom environments – particularly those that result in improved student outcomes – we recognize this does not control for instructors' fidelity of implementation. Fidelity of implementation is how well an intervention or activity is implemented in comparison with the original program's intention (O'Donnell, 2008; Stains & Vickrey, 2017), and this can strongly impact the efficacy of the pedagogy. So, as we continue to test active learning strategies, it is critical to describe how and why certain pedagogies are enacted in the classroom.

Third, we examined only three journals that commonly publish biology education research. This means the findings are not representative of *all* biology education research that has been published during that time period. However, the three journals we focused on are commonly used by the biology education research community. For example, Life Sciences Education and Journal of Microbiology and Biology Education publish primarily research articles and have a long-standing history and a large readership; *CourseSource* is the only online journal that exclusively publishes evidence-based biology teaching materials for undergraduate classrooms and laboratories.

Future work will (1) identify to what extent - and how - active learning is characterized *across* the DBER literature; (2) characterize the *definition* of active learning in the context of undergraduate STEM by collecting survey data from discipline-based education research (DBER) communities across STEM fields; (3) categorize the *specific active learning strategies* employed across STEM disciplines through survey data; and (4) investigate to what extent, if at all, perceptions of active learning differ among DBERs across STEM fields.

CONCLUSIONS

We support the use of active learning as a unifying term to generate awareness and collaboration among those interested in improving their teaching. The term gives DBER instructors an accessible on-ramp to engage with larger initiatives. However, because the term is rarely defined and can have many different meanings, those who use active learning should define what they mean and give examples of the strategies they are using. For example, authors could say: "We used an active learning instructional approach focused on student engagement using group work and clicker questions with peer instruction," followed by the appropriate citations and additional detail about the application and frequency of strategies. These additional details will allow the community to address more nuanced questions, such as: Do specific active learning instructional strategies promote student learning in multiple environments? Which strategies increase equitable outcomes for students from diverse backgrounds? and How can we maximize the effectiveness of particular active learning strategy in a variety of contexts? These questions can be more effectively answered when the approach and context of the learning environment is precisely defined. This clarity has the potential to make discipline-based education research communities, and their research, stronger.

ACKNOWLEDGMENTS

We are grateful to the discipline-based education research group at Auburn University for valuable feedback; to Taylor McKibben, Sara Wood, Brian Peters, and Brittany Woodruff for helping with data collection and analysis. We would also like to thank Doug Lombardi and Tim Shipley for their support and encouragement, and two anonymous reviewers for their critical insights that greatly improved the manuscript.

REFERENCES

Ackermann, E. (2001). Piaget's constructivism, Papert's constructionism: What's the difference. *Future of learning group publication*, *5*(3), 438.

Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., ... & Weber, N. (2019). *EDUCAUSE Horizon Report 2019 Higher Education Edition* (pp. 3-41). EDU19.

American Association for the Advancement of Science. (2009, July). Vision and change: A call to action. Retrieved January 29, 2020, from https://live-visionandchange.pantheonsite.io/wp-content/uploads/2011/03/VC-Brochure-V6-3.pdf

Aragón, O. R., Eddy, S. L., & Graham, M. J. (2018). 'Faculty Beliefs about Intelligence Are Related to the Adoption of Active-Learning Practices', *CBE—Life Sciences Education*, 17/3: ar47. Am Soc Cell Biol.

Auerbach, A. J., & Schussler, E. (2017). A Vision and Change Reform of Introductory Biology Shifts Faculty Perceptions and Use of Active Learning. *CBE—Life Sciences Education*, *16*(4), ar57.

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). 'Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning.', *CBE-Life Sciences Education*, 16/4: ar56.

Becker, E. A., Easlon, E. J., Potter, S. C., Guzman-Alvarez, A., Spear, J. M., Facciotti, M. T., ... & Pagliarulo, C. (2017). The effects of practice-based training on graduate teaching assistants' classroom practices. *CBE—Life Sciences Education*, *16*(4), ar58.

Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., ... Risely, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. Research-Based Reform of University Physics, 1(1), 2–39.

Bentley, M., and Connaughton, V.P. 2017. A simple way for students to visualize cellular respiration: Adapting the board game Mousetrap^{TM} to model complexity. CourseSource. https://doi.org/10.24918/cs.2017.8

Bouwma-Gearhart, J. L., Ivanovitch, J. D., Aster, E. M., & Bouwma, A. M. (2018). Exploring postsecondary biology educators' planning for teaching to advance meaningful education improvement initiatives. *CBE—Life Sciences Education*, *17*(3), ar37.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

Cavanagh, A. J., Chen, X., Bathgate, M., Frederick, J., Hanauer, D. I., & Graham, M. J. (2018). Trust, growth mindset, and student commitment to active learning in a college science course. *CBE—Life Sciences Education*, *17*(1), ar10.

Coffield, F. (2008). *Just suppose teaching and learning became the first priority...* London: Learning and Skills Network.

Connell, G. L., Donovan, D. A., & Chambers, T. G. (2016). Increasing the use of student-centered pedagogies from moderate to high improves student learning and attitudes about biology. *CBE—Life Sciences Education*, *15*(1), ar3.

Cook, B. G., Smith, G. J., & Tankersley, M. (2012). Evidence-based practices in education.

Cooper, K. M., & Brownell, S. E. (2016). Coming out in class: Challenges and benefits of active learning in a biology classroom for LGBTQIA students. *CBE—Life Sciences Education*, *15*(3), ar37.

Cooper, K. M., Ding, L., Stephens, M. D., Chi, M. T., & Brownell, S. E. (2018). A course-embedded comparison of instructor-generated videos of either an instructor alone or an instructor and a student. *CBE—Life Sciences Education*, *17*(2), ar31.

Cooper, K. M., Ashley, M., and Brownell, S. E. (2017). 'A bridge to active learning: A summer bridge program helps students maximize their active-learning experiences and the active-learning experiences of others.' *CBE—Life Sciences Education* 16.1, ar17.

Cooper, K. M., Haney, B., Krieg, A., & Brownell, S. E. (2017). What's in a name? The importance of students perceiving that an instructor knows their names in a high-enrollment biology classroom. *CBE—Life Sciences Education*, *16*(1), ar8.

Dewy, J. (1916). Democracy and education. *New York: McMillan*.

Dewsbury BM. (2017). Context determines strategies for "activating" the inclusive classroom. *J Microbiol Biol Educ.,* 18(3) doi: 10.1128/jmbe.v18i3.1347.

Drew, V., & Mackie, L. (2011). Extending the constructs of active learning: implications for teachers' pedagogy and practice. *Curriculum Journal*, *22*(4), 451-467.

Durham, M. F., Knight, J. K., & Couch, B. A. (2017). 'Measurement Instrument for Scientific Teaching (MIST): a tool to measure the frequencies of research-based teaching practices in undergraduate science courses', *CBE—Life Sciences Education*, 16/4: ar67. Am Soc Cell Biol.

Eddy, S L, Converse, M., & Wenderoth, M. P. (2015). 'PORTAAL: A Classroom Observation Tool Assessing Evidence-Based Teaching Practices for Active Learning in Large Science, Technology, Engineering, and Mathematics Classes', *CBE Life Sci Educ*, 14/2: 14:ar23. DOI: 10.1187/cbe.14-06-0095

Eddy, Sarah L, & Hogan, K. A. (2014). ‘Getting under the hood: how and for whom does increasing course structure work?’, *CBE-Life Sciences Education*, 13/3: 453–68.

Elliott, E. R., Reason, R. D., Coffman, C. R., Gangloff, E. J., Raker, J. R., Powell-Coffman, J. A., & Ogilvie, C. A. (2016). Improved student learning through a faculty learning community: How faculty collaboration transformed a large-enrollment course from lecture to student centered. *CBE—Life Sciences Education*, *15*(2), ar22.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). ‘Active learning increases student performance in science, engineering, and mathematics’, *Proceedings of the National Academy of Sciences USA*, 111/23: 8410–5. DOI: 10.1073/pnas.1319030111

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The discovery of grounded theory; strategies for qualitative research. *Nursing research*, *17*(4), 364.

Goff, E. E., Reindl, K. M., Johnson, C., McClean, P., Offerdahl, E. G., Schroeder, N. L., & White, A. R. (2017). Efficacy of a meiosis learning module developed for the virtual cell animation collection. *CBE—Life Sciences Education*, *16*(1), ar9.

Green, N. H., McMahon, D. G., & Brame, C. (2016). Using Online Active-Learning Techniques to Convey Time Compensated Sun Compass Orientation in the Eastern North American Monarch. *Journal of microbiology & biology education*, *17*(3), 430.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). ‘Increased structure and active learning reduce the achievement gap in introductory biology’, *Science*, 332/6034: 1213–6. DOI: 10.1126/science.1204820

Hoefnagels, M., and Taylor, M.S. (2016). “Boost your evolution IQ”: An evolution misconceptions game. *CourseSource*. https://doi.org/10.24918/cs.2016.12

Hora, M. T., Oleson, A., & Ferrare, J. J. (2013). ‘Teaching dimensions observation protocol (TDOP) user’s manual’, *Madison: Wisconsin Center for Education Research*.

Jeno, L. M., Raaheim, A., Kristensen, S. M., Kristensen, K. D., Hole, T. N., Haugland, M. J., & Mæland, S. (2017). The relative effect of team-based learning on motivation and learning: a self-determination theory perspective. *CBE—Life Sciences Education*, *16*(4), ar59.

Kreber, C., & Cranton, P. A. (2000). Exploring the scholarship of teaching. *The journal of higher education*, *71*(4), 476-495.

Kudish, P., Shores, R., McClung, A., Smulyan, L., Vallen, E. A., & Siwicki, K. K. (2016). Active learning outside the classroom: implementation and outcomes of peer-led team-learning workshops in introductory biology. *CBE—Life Sciences Education*, *15*(3), ar31.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Lee, C. J., Toven-Lindsey, B., Shapiro, C., Soh, M., Mazrouee, S., Levis-Fitzgerald, M., & Sanders, E. R. (2018). Error-discovery learning boosts student engagement and performance, while reducing student attrition in a bioinformatics course. *CBE—Life Sciences Education*, *17*(3), ar40.

Lund, T. J., & Stains, M. (2015). The importance of context: an exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education*, *2*(1), 13.

McCourt, J. S., Andrews, T. C., Knight, J. K., Merrill, J. E., Nehm, R. H., Pelletreau, K. N., ... & Lemons, P. P. (2017). What motivates biology instructors to engage and persist in teaching professional development?. *CBE—Life Sciences Education*, *16*(3), ar54.

McLean, J. L., & Suchman, E. L. (2016). Using magnets and classroom flipping to promote student engagement and learning about protein translation in a large microbiology class. *Journal of microbiology & biology education*, *17*(2), 288.

Miller, S., & Tanner, K. D. (2015). A portal into biology education: an annotated list of commonly encountered terms. *CBE—Life Sciences Education*, *14*(2), fe2.

Montessori, M. (1946). *Education for a new world* (Vol. 1). Kalakshetra.

Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2016). A "scientific diversity" intervention to reduce gender bias in a sample of life scientists. *CBE—Life Sciences Education*, *15*(3), ar29.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of educational research*, *78*(1), 33-84.

Olson, S., & Riordan, D. G. (2012). 'Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President', *Executive Office of the President*.

Papert, S. (1980). Mindstonns. *New York: Basic Rooks*, *607*.

Pesavento, T., Klein, J., Macasaet, D., Shorter, C., & Wagstaff, S. (2015, July 6). History & Context for Active Learning. Retrieved from https://wisc.pb.unizin.org/teachingwithtech/chapter/history-context-for-active-learning/#:~:text=The term "active learning" and,and MOOCs are filling today.

Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D., Fagen, A. P., et al. (2009). 'Summer institute to improve university science teaching', *Science*, 324/5926: 470–1. American Association for the Advancement of Science.

Piaget, J. (1932). 1965. *The moral judgment of the child*.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). 'The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices', *CBE-Life Sciences Education*, 12/4: 618–27.

Stains, M., & Vickrey, T. (2017). Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE—Life Sciences Education*, *16*(1), rm1.

Stoltzfus, J. R., & Librakin, J. (2016). Does the room matter? Active learning in traditional and enhanced lecture spaces. *CBE-Life Sciences Education*, 15(4), ar68.

Strauss, A. L. (1987). *Qualitative analysis for social scientists.* Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Tanner, K. D. (2013). Structure matters: twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE—Life Sciences Education*, *12*(3), 322-331.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., et al. (2020). 'Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math', *Proceedings of the National Academy of Sciences*, 117/12: 6476–83. National Acad Sciences.

Turkle, S., & Papert, S. (1990). Epistemological pluralism: Styles and voices within the computer culture. *Signs: Journal of women in culture and society*, *16*(1), 128-157.

Vygotsky, L. S. (1987). The development of scientific concepts in childhood. *The collected works of LS Vygotsky*, *1*, 167-241.

Wilton, M., Gonzalez-Niño, E., McPartlan, P., Terner, Z., Christoffersen, R. E., & Rothman, J. H. (2019). Improving Academic Performance, Belonging, and Retention through Increasing Structure of an Introductory Biology Course. *CBE—Life Sciences Education*, *18*(4), ar53.

Chapter 2: Group work and student performance in biology: A meta-analysis.

Emily P. Driessen[a,*], Alan E. Wilson[b], Sara Odom[a], Sara Beth Ramsey[a], Sara Wood[a], and Cissy J. Ballen[a]

[a]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

[b]School of Fisheries, Aquaculture, and Aquatic Science, Auburn University, Auburn, AL 36849, USA

[*]Corresponding author:

Emily P. Driessen

Email: epd0016@auburn.edu

Phone: 763-268-9023

Address: 1161 W. Samford Ave. Auburn, AL 36849

Manuscript submitted as a research article to BioScience where it is under review.

Keywords: active learning, STEM, post-secondary, cooperative learning, life sciences

ABSTRACT

Active learning is hailed as the pinnacle of teaching in STEM classrooms for its ability to maximize learning. While previous work broadly defined active learning as the antithesis of lecture, measuring the effects of all teaching strategies combined, we tested the isolated impacts of group work on student performance, using meta-analysis. We collected data from 34 articles representing 66 estimates calculated from performance outcomes of 18,494 students. Our overall estimate indicates the implementation of group work in biology classrooms increased student performance by 0.72 standard deviations. When applied to a grade scale based on real-world data, the magnitude of this change is roughly one letter grade. Moderator analyses revealed the increase in performance holds across all class sizes, class levels, and group sizes up to five students. These results demonstrate group work leads to impressive boosts in student performance and reveal the value of studying specific active learning strategies.

BACKGROUND

Despite increasing interest concerning the impacts of active learning on student performance, recommendations of specific, discrete strategies to use in classroom environments remain vague. Current approaches to testing the effects of active learning in Science, Technology, Engineering, and Mathematics (STEM) are largely based on studies comparing courses with at least some active learning to those featuring traditional, undisrupted lecture (Andrews et al., 2011; Ballen et al., 2017; Barral et al., 2018; Casper et al., 2019; Freeman et al., 2011, 2014; Haak et al., 2011; Theobald et al., 2020). While such studies are critical to our broad understanding of the impacts of evidence-based teaching, these approaches provide little insight into the mechanisms by which students respond to different teaching methods. For a complete understanding of the effective elements that define active learning, we need to consider not only its overall impacts, but also the individual impacts of isolated active learning strategies on student outcomes. Here, we fill a gap in the literature by utilizing meta-analytic techniques to quantitatively evaluate the effect of one common active learning strategy: group work.

Previous research showed that group work is one of the most frequently cited examples of active learning strategies according to post-secondary biology educators (Driessen et al., 2020), and it is frequently included in research demonstrating positive student outcomes (Carmichael, 2009; Chaplin, 2009; Daniel, 2016; Donovan et al., 2018; Gaudet et al., 2010; Johnson et al., 1998; Knight & Wood, 2005; Marbach-Ad et al., 2016; Springer et al., 1999; Weir et al., 2019; Yapici, 2016). For the purposes of this meta-analysis, we broadly defined group work as students working together during class time in groups of at least two (and smaller than the total class size). This definition allowed us to capture a wide breadth of studies related to group work, encompassing specific types of group work such as cooperative-learning (Johnson,

Johnson, and Smith, 1991), think-pair-share (Prahl (2017), and team-based learning (Michaelsen & Sweet, 2008).

We compared results of experiments that documented student performance in classes with at least some group work component versus classes without a group work component by analyzing 66 estimates taken from 34 studies from the published and unpublished literature. In total, the data represents performance outcomes from 18,494 students with an average of 280 students per estimate. Gathering available data from nearly a century of research, we investigated the following question: What is the effect of group work on academic performance?

MATERIALS AND METHODS

Experimental design

We followed the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for best practices in quantitative reviews (Page et al., 2021). To identify studies comparing student performance in undergraduate biology classes that incorporated at least some group work component to those classes without a group work component, we (1) searched six online databases (Appendix E, Table S1) using a variety of search terms (Appendix E, Box S1); (2) reviewed papers included in three previous meta-analyses on active learning and group work (Freeman et al., 2014; Johnson et al., 2000; Springer et al., 1999); and (3) included other relevant works already known to the authors. *A priori*, we searched for articles published from 1 January 1924 to 16 November 2020. We selected our earliest date because it was cited in a previous article (Johnson et al., 1998) that informed one of the previous meta-analyses on group work (Johnson et al., 2000). Two of the authors (E.D. and

S.B.R.) conducted a systematic search using the updated PRISMA checklist and flow diagram (Page et al., 2021).

Criteria for admission

We established the criteria for selecting publications at the onset of this study and did not subsequently alter them. Criteria included studies that (1) examined post-secondary students in a biological science class; (2) incorporated any type of group work (as defined by two or more students working or discussing topics together); (3) were conducted in a classroom; (4) were published or reported from 1 January 1924 to 16 November 2020; (5) reported, at a minimum, a mean measure of student performance before and after group work, or a mean measure for the classes partaking in group work and for the classes not receiving group work; and (6) used the same concept inventory, similar or the same exams, or class grade to evaluate the two classes. Of note, we did not include any studies that compared the effect of group work to the effect of one-on-one student to teacher instruction as we did not find that representative of many post-secondary teaching environments. Additionally, we did not include cross-over design studies, meaning, if the same students were exposed to both individual and group work and then compared to themselves, then we did not include the study in this meta-analysis. Finally, we did not include studies where the only form of group work took place during group exams because we were interested in the effect of group work during class instruction rather than during summative assessments.

Review process

Two researchers (E.D. and S.B.R) read the abstracts of papers returned by database searches as well as citations in three previous relevant meta-analyses (Freeman et al., 2014; Johnson et al., 2000; Springer et al., 1999) to determine eligibility. We excluded articles from the study if they did not meet one or more of the criteria for admission. In cases where it was unclear whether a study fit the scope of our project, we read the entirety of the paper to determine its suitability, excluding articles that did not meet our inclusion criteria. Finally, we were left with 34 papers from which to collect data. For details concerning the number of papers returned by each database, the number of papers that made it past the abstract screening process, the number of papers that made it past the final screening process, and more, see Figure 2.1. When available, we recorded data on class size, class level, group size, and research design type (Table 2.1).

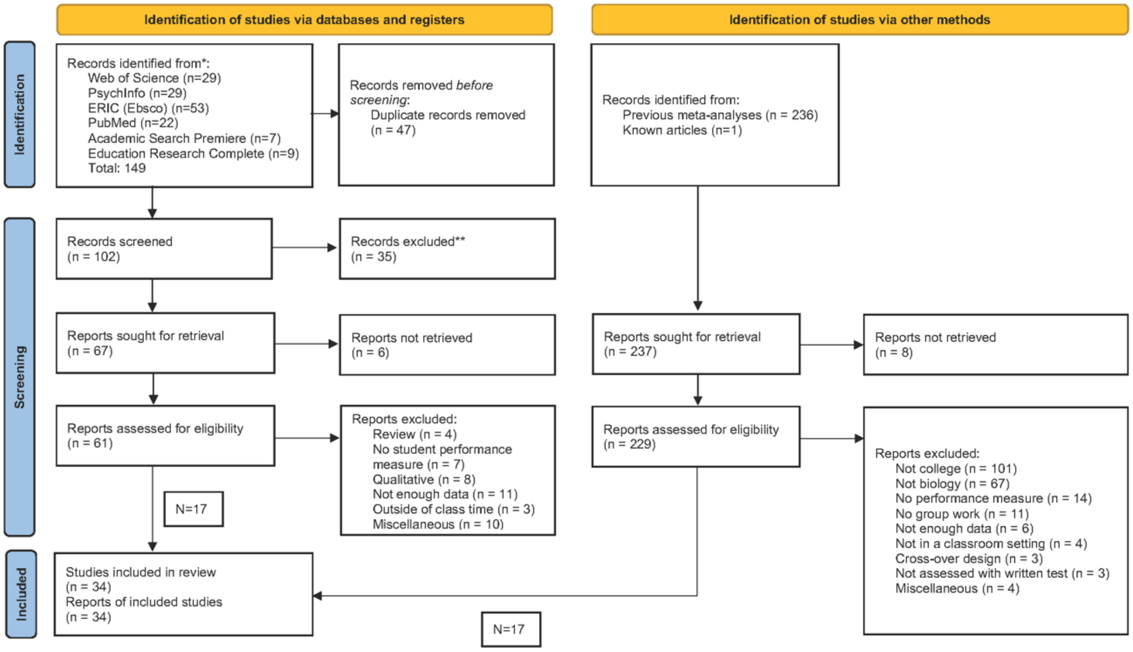


Figure 2.1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) literature analysis diagram.

Table 2.1 Studies included in the meta-analyses. Author and year columns are used to identify the study. The full citations are provided in the supplemental materials. The class size column details whether the class was small (S; less than 50 students), medium (M; 50-110 students), or large (L; 110+ students). The group size column details whether the group was small (S; two to four students), medium (M; four to five students), or large (L; more than five students). Lastly, the design column indicates whether a study had an independent (I) or paired (P) design.

Author(s)	Year	Class Size	Group Size	Class Level	Design
Altiparmak & Nakiboglutezer	2009	S	M	NA	I
Armbruster et al.	2009	L	M	L	I
Armstrong, Chang, & Brickman	2007	L	L	L	I
Armstrong, Chang, & Brickman	2007	L	L	L	I
Baines et al.	2004	S	S	L	I
Burrowes	2003	M	M	L	I
Carmichael	2009	L	M	L	I
Chaplin	2009	M	S	L	I
Chuck	2011	S	M	U	I
Co	2019	L	S	L	I
Costa Silva et al.	2018	S	M	NA	I
Daniel	2016	M	M	L	I
Das et al.	2019	S	L	G	I
Donovan, Connell, & Grunspan	2018	L	M	L	P
Donovan, Connell, & Grunspan	2018	L	M	L	P
Donovan, Connell, & Grunspan	2018	L	M	L	P
Fuller et al.	2016	L	S	NA	P
Gaudet et al.	2010	S	M	U	I
Hacisalihoglu et al.	2018	S	S	L	I
Huyusken et al.	2019	S	NA	L	I
Huyusken et al.	2019	S	NA	L	I
Jensen & Finley	1996	S	S	L	I
Jensen & Finley	1996	S	S	L	I
Kitchen et al.	2003	L	S	U	P
Klegeris, Bahniwal & Hurren	2013	M	L	U	P
Knight & Wood	2005	M	S	U	I
Lord	1997	M	M	L	I
Marbach-Ad et al.	2016	L	S	L	I
McCormick	2000	S	NA	L	I
McCormick	2000	S	NA	L	P
Montelone, Rintoul & Williams	2008	M	S	L	I
Nelson et al.	2009	L	S	L	I
Premo et al.	2018	S	M	L	I
Randolph	1992	S	M	L	I
Reich & Wang	2019	M	S	L	I
Theobald et al.	2017	L	S	L	P
Theobald et al.	2017	L	S	L	P
Walker et al.	2008	L	NA	L	I

Weasel & Finkel	2016	L	L	L	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	S	NA	U	P
Weir et al.	2019	M	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Weir et al.	2019	L	NA	NA	P
Yapici	2016	S	M	NA	I

Effect size

To account for differences in student performance (e.g., average class performance, final exam performance) across two groups, we calculated a standardized mean difference for each course in the form of Hedges' *g* (Hedges, 1981). This statistic is commonly used in meta-analyses, and it estimates the effect size for the difference between means scaled to the pooled standard deviation, but also includes a correction for small sample sizes (Hedges & Olkin, 1985). A positive value indicates increased student performance in the intervention group (i.e., group work) compared to the control group (i.e., no group work), while a negative value indicates increased student performance in the control group (i.e., no group work) compared to the intervention group (i.e., group work; Durlak, 2009).

Data collection

We categorized study designs as either: (i) paired (i.e., pre- versus post-group-work data collection) or (ii) independent (i.e., post-treatment with group work versus post-control without group work). The difference in study design mandated a difference in data collection. Specifically, for a paired design, we collected mean student performance scores (i.e., class performance, exam performance, or concept inventory performance), standard deviation, and the number of students in each class both before and after experiencing group work as well as a pre-post correlation value. However, correlation values were not present in any of the studies, so we inputted correlation values of 0.9 (Appendix E). On the other hand, for an independent design, we collected mean performance scores, standard deviation, and the number of student participants in each class (e.g., classes that experienced group work versus classes that did not experience group work) from each article. Despite the differences between the data collected for each experimental design type, we extracted data from both types of studies directly from tables or text in the publication or from figures using WebPlotDigitizer version 4.2 (https://automeris.io/WebPlotDigitizer). In cases where a study did not provide standard deviation/standard error values and/or the number of students in each group – the case for 8 of our 66 estimates (12.5%) – we estimated these values using the average standard deviation or average number of students from the other studies (mean-value imputation; Kambach et al., 2020) (Appendix E). We developed a protocol to handle cases where studies provided either multiple treatments, multiple controls, or multiple outcomes (Appendix E).

Coding Data

After data collection, we categorized each study according to their class size, class level, and group size. First, we coded studies based on whether they took place in small (less than 50 students), medium (50-110 students) or large (111+ students) classes. These codes were based on groupings from a previous meta-analysis on the effect of active learning on student performance (Freeman et al., 2014). Second, we categorized classes into lower-level (majority first- and second-year student classes) and upper-level (majority third- and fourth-year student classes) categories. Finally, we coded data based on group size, or number of students in a group. We coded group size as small (less than 4 students), medium (4-5 students), and large (more than 5 students) groups because these were the three natural clusters revealed by a histogram of the data (Appendix E).

Statistical analysis

We collected and analyzed data from two types of studies: (i) paired design; and (ii) independent design. Specifically, we used Comprehensive Meta-Analysis (CMA) software (https://www.meta-analysis.com/) to calculate Hedges' *g* and the associated variance for each study. We analyzed the Hedges' *g* values and associated variance values in R studio ((Team, 2016)), using the *metafor* package in R to conduct a random-effects meta-analysis fitted with the restricted maximum-likelihood method (REML) to find the overall effect of group work on student performance (Viechtbauer, 2010).

We then contextualized this overall effect by applying it to a real-world dataset consisting of final grades from students enrolled in fourteen large, introductory biology courses at a single institution. The average number of students in each class was 211 with a total of 2,951 students participating across all 14 sections. We multiplied the average standard deviation for each class section (i.e., 13.49) by the overall estimate to calculate how group work would affect a student's grade.

Next, we assessed this overall finding for sensitivity and bias using *metafor* (Viechtbauer, 2010). Specifically, to test the sensitivity of our results, we conducted meta-analyses with the following changes from the original model: (1) we removed the random effect of author from the original analysis; (2) we excluded one study that contributed 26 of the 66 estimates and re-ran the meta-analysis to compare the overall effects; (3) we removed an extreme outlier (i.e., ~25 standard errors larger than the mean; Yapici, 2016); (4) we doubled the imputed standard deviation value for those studies missing standard deviations; we examined the data by the two types of research designs we considered in our meta-analysis by conducting a pairwise comparison; (5) we tested whether our dataset was robust to changes in the correlation value in paired research design comparisons by changing it to 0.5 and then 0.1; and (6) we reanalyzed the dataset after excluding data from paired design studies. To assess publication bias, we constructed and visually inspected them (Sterne et al., 2005), utilized Duval and Tweedie's trim and fill method (Duval & Tweedie, 2000a; 2000b), and calculated fail-safe Ns (Rosenthal, 1979).

Next, to assess whether the effect of group work on student performance, as compared to the effect of no group work on student performance, held across each of the moderating groups, we conducted mixed-model analyses with the restricted maximum-likelihood method (REML) for class size, class level, and group size as individual fixed effects, using the *metafor* package in R studio (Team, 2016; Viechtbauer, 2010). To visualize the output as orchard plots, we used the orchaRd package in R (Nakagawa et al., 2020). We then exported these plots to edit in Biorender (BioRender.com).

Finally, we conducted pairwise analyses with the restricted maximum-likelihood method (REML) to test for differences between the moderating sub-groups for class size, class level, and group size, using the *metafor* package in R studio (Team, 2016; Viechtbauer, 2010).

Data Availability

All data and code used for this meta-analysis are available here: https://github.com/EmilyDriessen/Group-work-and-student-performance-in-biology-A-meta-analysis..git

RESULTS

Overall effect size

We used a random effects model, with author as a random effect, to compare the effect of group work on student performance to the effect of no group work on student performance by using the difference between two means as divided by their pooled standard deviation (Hedges' *g*). Our results showed group work had a large positive effect on student academic performance (Hedges' $g = 0.72 \pm 0.26$ for ± 95% confidence intervals (CI), $Z = 5.38$, $P << 0.0001$; Figure 2.2A; Table S2A). This means, on average, students who worked in groups performed 0.72 standard deviations higher than students who did not work in groups. To put the effect size in perspective, Hedges' *g* values of 0.20 or higher on performance measures in education research should be of interest to policy makers (Hedges & Hedberg, 2007).

To contextualize our overall estimate of 0.72, we applied it to a real-world dataset by collecting final grades from students enrolled in fourteen large, introductory biology courses at a single institution. The average number of students in each class was 211 with a total of 2,951 students participating across all 14 sections. We multiplied the average standard deviation for each class section (i.e., 13.49) by the estimate (i.e., 0.72) to calculate how group work would affect a student's grade. This yielded an increase of 9.71 percentage points, roughly equal to one letter grade.

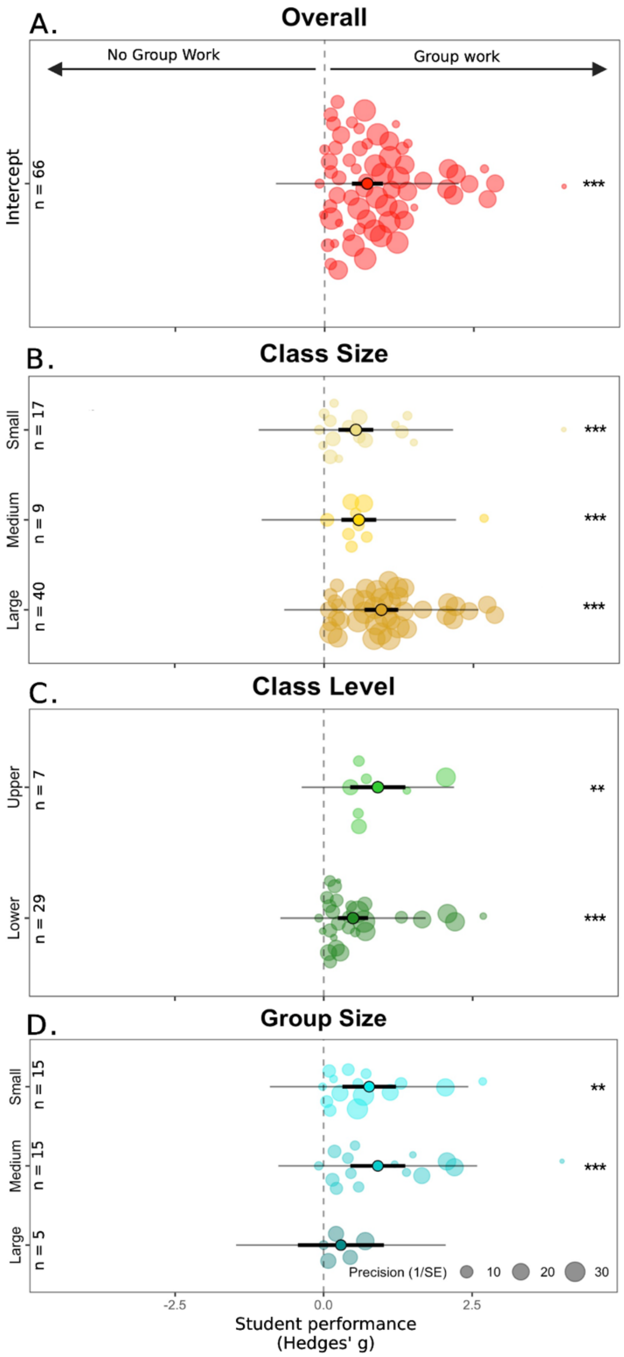


Figure 2.2 Orchard Plot depicting the effect of group work on student performance.

Orchard plots demonstrate the impact of group work on student performance using Hedges' g. Each individual circle represents one comparison of the effect of group work on student performance versus the effect of no group work on student performance, and the size of each of these comparisons is inversely proportional to its standard error. Dark horizontal lines represent 95% confidence intervals on either side of the mean, and the lighter lines represent the prediction interval. If the estimate, including the 95% confidence intervals, crosses zero, then the effect of group work is not significantly different from the effect of no group work on student performance. **A**) All comparisons. **B**) Comparisons as parsed by class size (small = less than 50 students; medium = 50 to 110 students; large = more than 110 students). **C**) Comparisons as parsed by class level (lower = first- and second-year students; upper = third- and fourth-year students). **D**) Comparisons as parsed by group size (small = two to four students; medium = four to five students; large = more than five students).

We then conducted several sensitivity analyses to determine the robustness of our results by considering the extent to which they are affected by changes in our models or assumptions (SI Results). Sensitivity analyses demonstrated the effect of group work on student performance is still statistically significant and robust to the removal of particularly large estimates, the removal of articles that contribute a large number of estimates, inputted standard deviation changes, correlation value changes, and the removal of estimates taken paired design studies (Table S2B-H). Additionally, unpublished studies with low effect sizes did not create sampling bias, as demonstrated by both Duval and Tweedie’s trim and fill method (Duval & Tweedie, 2000a; Duval & Tweedie, 2000b) and Rosenthal’s fail-safe N method (Rosenthal, 1979)(Appendix F; Figure S1).

Moderator analyses

After conducting an overall meta-analysis, in addition to sensitivity and bias analyses, we analyzed our data by class size, class level, and group size, to assess whether the positive and statistically significant effect of group work on student performance, as compared to the effect of no group work on student performance, held across each of the moderating groups. We found that group work improved student performance across all class sizes (i.e., small, medium, large; Figure 2.2B; Table 2.2; QM (df = 3) = 103.82, P < 0.0001), though the largest positive effect of group work was in large classes (i.e., more than 110 students). Across both lower- and upper-level classes, group work had a significant effect on student performance (Figure 2.2C; Table 2.3; QM (df=2) = 29.05, P < 0.0001). However, we observed larger effects of group work among upper-level classes (i.e., third- and fourth-year student classes). Lastly, we discovered group work had a significant effect on student performance in small (< 4 students) and medium group sizes (4-5 students) (Figure 2.2D; Table 2.4; QM (df=3) =26.65, P < 0.0001). However, we did not observe a significant effect of group work on the performance of students assigned to groups with more than 5 people (i.e., large groups) although the same size for this group was small (n = 5). These analyses reveal that the significant and positive effect of group work on student performance holds across all class sizes, class levels, and group sizes of five or fewer students.

Table 2.2 The effect of group work on student performance, as moderated by class size.

Class size	***n****	**Hedges' *g***	**s.e.**	**p-val**	**95% C.I.: lower limit**	**95% C.I.: upper limit**
Small (≤ 50)	17	0.5320	0.1492	0.0004	0.2396	0.8244
Medium (50 < n ≤ 110)	9	0.5812	0.1492	<0.0001	0.2888	0.8735
Large (> 110)	40	0.9584	0.1439	<0.0001	0.6763	1.2404

*Note that *n*'s may not sum to 66 due to missing data.

Table 2.3 The effect of group work on student performance, as moderated by class level.

Class Level	***n****	**Hedges' *g***	**s.e.**	**p-val**	**95% C.I.: lower limit**	**95% C.I.: upper limit**
Graduate**	1	0.0000	0.6419	1.0000	-1.2580	1.2580
Upper	7	0.9074	0.2364	0.0001	0.4440	1.3708
Lower	29	0.4898	0.1294	0.0002	0.2361	0.7435

*Note that *n*'s may not sum to 66 due to missing data.

**Additionally, note that the group "graduate" is not represented in Figure 2.2. This is because there was only one comparison for this subgroup. For this reason, we left it out, but felt it important for transparency to include the data here.

Table 2.4 The effect of group work on student performance, as moderated by group size.

Group size	***n****	**Hedges' *g***	**s.e.**	**p-val**	**95% C.I.: lower limit**	**95% C.I.: upper limit**
Small (less than 4)	15	0.7659	0.2305	0.0009	0.3141	1.2178
Medium (4-5)	15	0.9120	0.2355	0.0001	0.4504	1.3736
Large (5+)	5	0.2905	0.3699	0.4322	-0.4344	1.0154

*Note that *n*'s may not sum to 66 due to missing data.

Pairwise analyses of moderators

After testing the effect of group work on student performance, compared to the effect of no group work on student performance, for each moderating group, we conducted pairwise analyses to test for differences between those moderating groups. Based on these analyses, we observed the effect of group work varied based on class size (QM (df=2) = 77.73, P<0.0001), with students from large classes performing 0.43 standard deviations better than students from small classes (+/-0.15; +/- 95% C.I.; p<0.0001) and 0.38 standard deviations better than students in medium class sizes (+/- 0.11; +/- 95% C.I.; p<0.0001). Students in medium class sizes performed 0.05 standard deviations (+/- 0.18; +/- 95% C.I.) better than students in small class sizes, although this was not statistically significant (p=0.54). While there were differences between class size groups, there were no statistical differences on the impact of group work on performance outcomes between the class levels (QM (df=1) = 2.40, P=0.12) or the group size (QM (df=2) = 2.02, P=0.36).

DISCUSSION

Using meta-analysis, we show the implementation of group work increased overall student performance by 0.72 standard deviations, approximately an entire letter grade based on real student data, which is a greater effect than previous studies evaluating the effect of all active learning strategies combined (Freeman et al., 2014). Our results demonstrate group work increases performance in all class sizes, undergraduate class levels, and group sizes of five or less students. To evaluate the generality of impacts of active learning on student outcomes, we need research on a wide range of individual strategies. Future research will also profit from addressing how different applications of the same strategy affect students (Century et al., 2010).

Instructor Recommendations

Our work demonstrates the significant and large effect of group work on student performance, and we encourage instructors to implement group work in their biology classrooms. There are plenty of resources available for instructors who are interested in implementing group work, in its various forms, in their class. For example, Tanner, Chatman, and Allen (2017) detail several essential elements of effective cooperative learning (a specific type of group work), such as individual and group accountability and promoting interpersonal skills. Prahl (2017) advises instructors who wish to implement think-pair-share exercises in their classroom, offering guidance on writing questions and problems, interacting with the students, and assessing the exercises. Michaelsen and Sweet (2008) explain team-based learning and define the four essential elements as properly formed and managed groups, student accountability for the quality of their individual and group work, frequent and timely feedback on student work, and group assignments that promote learning and team development. Finally, Smith and Knight (2020) detail how clickers (technology used to relay student responses to the instructor) can be integrated into the biology classroom to help encourage students to apply their knowledge and analyze new scenarios. Such methods can easily be used in groups to allow students to think through questions together before they respond with their clickers. Although there are many other resources available for instructors interested in implementing group work in their class, we offer these as good starting points.

Limitations

The present study was limited in two ways. First, the studies we analyzed usually used a variety of group-work oriented strategies (e.g., group brainstorming, group discussion of homework questions, and group clicker questions). Because we cannot disentangle each specific group-work strategy, we cannot identify exactly which group-work-oriented strategy produced the largest effect on student performance. Identifying the best group work strategy for increasing student performance is important in making instructor recommendations, and this requires further study. Second, some of the studies included in this meta-analysis lacked information for all three of the moderators we analyzed (i.e., class size, class level, and group size). Missing information means we rely on subsets of data to evaluate the effects of moderators on group work. Conducting follow-up studies of class size, class level, and group size could certainly provide additional insights.

IMPLICATIONS

From an institutional perspective, we demonstrate the economic importance of testing and implementing evidence-based teaching strategies. State funding is responsible for 54% of public institution educational revenue (Laderman, 2017), and this funding is based on the performance of a university or college (e.g., student retention, graduation, transfer, and job placement outcomes; Dougherty et al., 2014). We know students generally earn lower grades in STEM courses relative to non-STEM courses (Koester et al., 2016), and these grades are consequential: those who leave STEM cite the performance challenge associated with STEM courses as primary motivation for their decision (Seymour & Hewitt, 1997; Seymour & Hunter, 2019). Thus, teaching strategies, such as group work, that promote student performance will benefit individuals as well as institutions that are evaluated based on aggregate outcomes.

ACKNOWLEDGEMENTS

We thank Elli Theobald, Abby Beatty, and Todd Steury for their invaluable advice. Additionally, we thank Jordan Harshman, Karen McNeal, and Mary Mendonca, the Ballen lab, and the entire Auburn University Biological Sciences department for their thoughtful feedback. Finally, we thank the many authors of research articles on the effect of group work who provided missing data upon request. This work was supported by NSF DUE-2120934, DUE-2011995, and DBI-1919462 awarded to CJB.

REFERENCES

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE Life Sciences Education*, *10*(4), 394–405. https://doi.org/10.1187/cbe.11-07-0061

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education*, *16*(4), ar56.

Barral, A. M., Ardi-Pastores, V. C., & Simmons, R. E. (2018). Student learning in an accelerated introductory biology course is significantly enhanced by a flippedlearning environment. *CBE Life Sciences Education*, *17*(3), 1–9. https://doi.org/10.1187/cbe.17-07-0129

Carmichael, J. (2009). Team-based learning enhances performance in introductory biology. *Journal of College Science Teaching*, *38*, 54–61. http://login.ezproxy.library.ualberta.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ofs&AN=508046050&site=ehost-live&scope=site

Casper, A. M., Eddy, S. L., & Freeman, S. (2019). True Grit: Passion and persistence make an innovative course design work. *PLoS Biology*, *17*(7), 1–10. https://doi.org/10.1371/journal.pbio.3000359

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218.

Chaplin, S. (2009). Assessment of the impact of case studies on student learning gains in an introductory biology course. *Journal of College Science Teaching*, *39*(1), 72.

Daniel, K. L. (2016). Impacts of Active Learning on Student Outcomes in Large-Lecture

Biology Courses. *The American Biology Teacher*, *78*(8), 651–655. https://doi.org/10.1525/abt.2016.78.8.651

Donovan, D. A., Connell, G. L., & Grunspan, D. Z. (2018). Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. *CBE Life Sciences Education*, *17*(4), 1–14. https://doi.org/10.1187/cbe.17-12-0283

Dougherty, K. J., Jones, S. M., Lahr, H., Natow, R. S., Pheatt, L., & Reddy, V. (2014). Performance funding for higher education: Forms, origins, impacts, and futures. *The ANNALS of the American Academy of Political and Social Science*, *655*(1), 163–184.

Driessen, E. P., Knight, J. K., Smith, M. K., & Ballen, C. J. (2020). Demystifying the meaning of active learning in postsecondary biology education. *CBE—Life Sciences Education*, *19*(4), ar52.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928.

Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

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

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE Life Sciences Education*, *10*(2), 175–186. https://doi.org/10.1187/cbe.10-08-0105

Gaudet, A. D., Ramer, L. M., Nakonechny, J., Cragg, J. J., & Ramer, M. S. (2010). Small-group learning in an upper-level university biology class enhances academic performance and student attitudes toward group work. *PLoS ONE*, *5*(12). https://doi.org/10.1371/journal.pone.0015821

Haak, D., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). *Increased structure and active learning reduce the achievment gap in introductory biology*. *June*, 1213–1216.

Hedges, L. V, & Olkin, I. (1985). *Statistical Methods for Meta-Analysis. Orlando, FL: Academic Pres*. Inc.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, *60*(7), 627–636.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative Learning Returns To College What Evidence Is There That It Works? *Change: The Magazine of Higher Learning*, *30*(4), 26–35. https://doi.org/10.1080/00091389809602629

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis.

Kambach, S., Bruelheide, H., Gerstner, K., Gurevitch, J., Beckmann, M., & Seppelt, R. (2020). Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecology and Evolution*, *10*(20), 11699–11712.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(4), 298–310.

Koester, B. P., Grom, G., & McKay, T. A. (2016). Patterns of gendered performance difference in introductory STEM courses. *ArXiv Preprint ArXiv:1608.07565*.

Laderman, S. (2017). State Higher Education Finance FY 2016. *State Higher Education Executive Officers*.

Marbach-Ad, G., Rietschel, C., & Thompson, K. V. (2016). Validation and application of the survey of teaching beliefs and practices for undergraduates (STEP-U): Identifying factors associated with valuing important workplace skills among biology students. *CBE Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-05-0164.

Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. *New directions for teaching and learning*, *2008*(116), 7-27.

Nakagawa, S., Lagisz, M., O'Dea, R. E., Rutkowska, J., Yang, Y., Noble, D. W. A., & Senior, A. M. (2020). *orchaRd: An R package for drawing 'orchard' plots (and 'caterpillars' plots) from meta-analyses and meta-regressions with categorical moderators*.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*.

Prahl, K. (2017). Best practices for the think-pair-share active-learning technique. *The American Biology Teacher*, *79*(1), 3-8.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving*. Westview Press, Boulder, CO.

Seymour, E., & Hunter, A.-B. (2019). Talking about leaving revisited. *Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*.

Smith, M. K., & Knight, J. K. (2020). Clickers in the biology classroom: Strategies for writing and effectively implementing clicker questions that maximize student learning. In *Active Learning in College Science* (pp. 141-158). Springer, Cham.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21–51. https://doi.org/10.3102/00346543069001021

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 75–98.

Tanner, K., Chatman, L. S., & Allen, D. (2003). Approaches to cell biology teaching: cooperative learning in the science classroom—beyond students working in groups. *Cell biology education*, *2*(1), 1-5.

Team, Rs. (2016). *RStudio: integrated development for R. RStudio, Inc., Boston, Massachusetts, USA*.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Nicole Arroyo, E., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard university press.

Weir, L. K., Barker, M. K., McDonnell, L. M., Schimpf, N. G., Rodela, T. M., & Schulte, P. M. (2019). Small changes, big gains: A curriculum-wide study of teaching practices and student learning in undergraduate biology. *PLoS ONE*, *14*(8), 1–16. https://doi.org/10.1371/journal.pone.0220900

Yapici, İ. Ü. (2016). Effectiveness of Blended Cooperative Learning Environment in Biology Teaching: Classroom Community Sense, Academic Achievement and Satisfaction. *Journal of Education and Training Studies*, *4*(4), 269–280. https://doi.org/10.11114/jets.v4i4.1372

Chapter 3: Evaluating open-note exams: student perceptions and preparation methods in an undergraduate biology class

Emily P. Driessen, Abby E. Beatty & Cissy J. Ballen

Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

This manuscript was accepted and published as a research article in PLOS ONE in 2022.

Keywords: study habits, performance, student affect, application, life sciences

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the GitHub repository, https://github.com/aeb0084/Evaluating-Open-Note-Examinations.git

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceptualization: EPD, CJB

Methodology: EPD, CJB

Validation: EPD

Formal Analysis: EPD, AEB

Investigation: EPD

Resources: EPD

Data Curation: EPD, AEB

Writing – Original Draft: EPD, AEB, CJB

Writing – Review & Editing: EPD, AEB, CJB

Visualization: EPD & AEB

Supervision: EPD, CJB

Project Administration: EPD

ABSTRACT

Although closed-note exams have traditionally been used to evaluate students in undergraduate biology classes, open-note exams are becoming increasingly common, though little is known about how students prepare for these types of exams. We investigated student perceptions of and their preparation habits for online open-note exams in an undergraduate biology class, as compared to their previous experiences with closed-note exams in other classes. Specifically, we explored the following research questions: (1a) How do students perceive open-note exams impact their exam scores, their anxiety, the amount they studied, and the amount their peers studied? (1b) How do these perceptions impact performance outcomes? (2a) How do students prepare for open-note exams? (2b) How do these preparation methods impact to performance outcomes? Results demonstrate students perceived increased exam scores, decreased exam-anxiety, decreased study time spent personally, and decreased study time spent by their peers for open-note exams, as compared to past experiences with closed-note exams. Open-ended survey responses analyzed through first- and second-cycle analyses showed students adapted their study habits by focusing on note preparation and broad conceptual understanding rather than rote memorization. Using linear mixed effects models to assess student performance, we found students who focused on understanding, note preparation and using external resources outperformed students who did not report those study habits. We conclude that, as institutions shift towards flexible and scalable assessments that can be used in face-to-face or online environments, the use of open-note exams can promote effective study habits and reward higher-order thinking with intentional guidance from the instructor.

BACKGROUND

Calls to action make specific recommendations for improving undergraduate biology education nationwide. For example, the American Association for the Advancement of Science (2009) outlined several core competencies intended to guide undergraduate biology education, including applying the process of science and understanding the relationship between science and society. Consistently and notably absent from calls for change is the recommendation for students' rote memorization of biology vocabulary or facts.

Nevertheless, one way in which student competencies are evaluated in higher education (i.e., Low-level Bloom's Taxonomy closed-note multiple-choice question exams; Momsen et al., 2010) rewards simple recall-based memorization of details rather than the development of the diverse skills needed to succeed in the scientific workforce (Momsen et al., 2010; Seymour & Hewitt, 1997; Wood, 2009; Zheng et al., 2008). These types of closed-note exams are scalable, easily assessed, and perceived by some as more rigorous. However, open-note exams – those where students can consult textbooks, notes, or other class-related material during the exam – can be designed to be scalable and easily-assessed as well, and, *in addition,* open-note exams can reward a different skillset – one more aligned with recommendations for improving undergraduate biology education.

Contrary to closed-note exams, open-note exams render basic recall questions useless because students can simply refer to their notes. Rather, instructors may adapt open-note exams by developing assessments that evaluate higher-level skills such as conceptualization, problem solving, and reasoning (Feller, 1994). For this reason, they have the potential to shift exam norms (Momsen et al., 2013), as other disciplines have demonstrated (Eilertsen & Valdermo, 2000). However, the literature concerning open-note exams is mixed, producing both proponents and opponents of open-note exams.

Proponents of open-note exams propose a variety of benefits. First, proponents of open-note exams claim students learn how to gather and critically analyze material from multiple sources, as opposed to closed-note exams which reward short-term storage and quick retrieval (Ambrose et al., 2010; Krasne et al., 2006; Theophilides & Koutselini, 2000). Additionally, open-note exams decrease students' exam anxiety (Block, 2011; Gharib, Phillips & Mathew, 2012; Williams and Wong, 2007) and minimize a desire to cheat (King, Guyette & Piotrowski, 2009; Nguygen et al., 2020; Watson & Sottile, 2010). Green, Ferrante, and Heppard (2016) suggested both that (1) higher education must evolve along with culture and technology, and (2) the ability to answer fact-based multiple-choice questions by rote memory is not adequately preparing students for future careers. Rather, instructional approaches that foster innovation, creativity, and independent thinking, such as open-note exams, train students for real-world operational decision-making (Green, Ferrante & Heppard, 2016).

Opponents of open-note exams suggest a few concerns. First, they fear open-note exams lull students into a false sense of grade security, raising students' *predictions* about their grades while not actually increasing student performance overall (Ioannidou, 1997; Jensen & Moore, 2009; Sato et al., 2015). However, these claims have been contested, as other work shows open-note exams increased performance, as compared to closed-note exams (Agarwal et al., 2008; Agarwal & Roediger, 2011). Jensen and Moore (2009) suggest that an inflated perception of one's grades may encourage negative behaviors or study habits such as lower levels of class attendance and engagement, and less time spent reviewing class materials. Moore & Jensen (2007) showed that compared to a class section who took closed-note exams, students assigned to a section that completed open-note exams were less likely to attend pre-exam help sessions, submit extra-credit assignments, and attend lecture.

Despite the mixed findings concerning open-note exams, research on open-note exams has continued, analyzing more nuanced scenarios. For example, Sato et al., (2015) explored the effect of open-note versus closed-note exams on student performance in the context of a biology classroom, paying particular attention to performance by Bloom's levels. They found little difference in student performance on questions, regardless of their Bloom's levels, between open- and closed-note exam takers. Sato et al. (2015) were particularly surprised that students who could use notes on Bloom's 1 and 2 questions (i.e., memorization-based questions) did not outperform students who could not use notes. They concluded that to understand the relationship between open-note exams, performance, and learning, future research must focus on how students adapt their study habits to open-note exams.

Study habits include strategies students use to understand and retain class content. They also encompass how much time students spend studying and how students distribute their study time over the class of a semester (Blaisman, Dunlosky, & Rawson, 2017; Walck-Shannon, Rowell, & Frey, 2021). Because exam performance is highly correlated with study habits (Alpert, Couch & Harmon, 2016: Bettinger et al., 2017; Xu & Jaggars, 2011; Xu & Jaggars, 2014), the effectiveness of open-note exam taking may rely heavily on the specific exam preparation methods used by students. For this reason, we explored the impact of study habits for open-note exams on student performance. We focused on student performance because high-stakes exams often account for a sizeable proportion of students' grades and can determine whether a student continues in STEM or leaves the field altogether (Seymour et al., 2019).

Additionally, student perceptions of open-note exams may moderate student performance. For example, if students predict their grades will be higher simply by being allowed to use their notes on exams, then students may believe they do not need to study as much as they would for a closed-note exam (Ioannidou, 1997; Jensen & Moore, 2009; Sato et al., 2015). This misled belief might impact preparation methods and consequently performance (Walck-Shannon, Rowell, & Frey, 2021; Xu et al., 2021). For this reason, we correlated the amount and direction that students perceived their exam score would change for open-note exams compared to closed-note exams with performance. We also analyzed the relationship between the amount of time an individual studied for open-note exams compared to closed-note exams and performance, as well as the amount of time an individual perceived their peers studied for open-note exams. Another student perception we examined was test anxiety because of previous research demonstrating the strong impact of anxiety on student performance in college courses (Ballen, Salehi, & Cotner, 2017; Cassady & Johnson, 2002; Chapell et al., 2005; England, Brigati, & Schussler, 2017; England et al., 2019; Ewell, Josefson, & Ballen, 2022; Salehi et al., 2019; Szafranski, Barrera, & Norton, 2012; Vitasari et al., 2010; Zeidner, 1998).

In this study, we examined study habits for and student perceptions of open-note exams longitudinally, over the course of one semester, surveying students after each of three mid-term open-note exams. Our longitudinal design is purposeful, given that previous literature shows student study habits change over time, after initial exposure to a new exam format. For example, Sato et al. (2015) examined whether the impact of open-note testing varied based on the extent of exposure to open-note exams, comparing those students who received open-note exams over the course of the entire quarter versus those with no prior open-note exam experience. Findings demonstrated differences in both perceptions and performance between the two groups, suggesting that students adjust how they prepare for open-note exams after repeated exposure to the assessment format, highlighting the importance of a longitudinal examination of student perceptions and behaviors.

Open-note exams have surged in popularity alongside nation-wide increases in online classwork (Hussar et al., 2020) and increased effort toward equitable and inclusive teaching strategies. However, to our knowledge, no study has addressed *how students study* differently for open-note exams compared to closed-note exams in undergraduate biology classes, if at all, and how different approaches relate to performance outcomes. Additionally, the correlation between student *perceptions of* open-note exams and student performance has yet to be examined.

To address these gaps in the literature, we investigated student perceptions of and their preparation habits for open-note exams in an undergraduate biology class, as compared to their previous experiences with closed-note exams in other classes. Specifically, we investigated the following research questions: (1a) How do students perceive open-note exams impact their exam scores, their anxiety, the amount they studied, and the amount their peers studied? (1b) How do these perceptions impact performance outcomes? (2a) According to open-ended responses, how do students prepare for open-note exams? (2b) How do these preparation methods impact performance outcomes? To address our research questions, we surveyed students after three open-note exams over a single semester, documented their perceptions and study habits, and analyzed how their responses related to performance outcomes.

METHODS

Research Design

This study used a mixed-methods approach, combining both qualitative and quantitative research methods and analyses. Our goal was to elucidate student perceptions of and study habits for open-note exams by collecting Likert scale (quantitative data) and open-response (qualitative data) survey data. We then related these two data types to student exam performance at three time points (i.e., Exam 1, Exam 2, and Exam 3) in the course using linear mixed-effects models. This research design allows us to explore the impact of (1) student perceptions of open-note exams on exam performance and (2) student study habits for open-note exams on exam performance. By using a repeated measures design (i.e., collecting survey responses after each of the three mid-term exams), we captured collective snapshots of our students' perceptions of and study habits for open-note exams over the semester. Of note, Auburn University's Institutional Review Board determined our research exempt and gave us written consent to proceed (protocol #20-603 EX 2012).

Participants

Students in this study were enrolled in an online, introductory-level undergraduate biology course, Organismal Biology, taught during the spring 2021 semester by two of the authors (EPD & CJB). This course is the second in an introductory biology sequence, designed for those in biology-related majors to prepare them for future classwork. The study took place in a primarily white research-intensive university in the southeastern region of the United States.

Course Description

The Organismal Biology course is a three-credit class that includes two 75-min class sessions each week and is offered every semester. The main objective of the class is for students to develop an understanding of eukaryotic evolution, classification, structure, and the spectacular diversity of living organisms. In this class, students were randomly assigned to groups (ranging from 5 to 7 members) prior to the first class period, and these group assignments lasted for the duration of the semester. This class was online due to the COVID-19 pandemic, so students met in their groups in breakout rooms in Zoom during each class period. Each class period included lecture and group activities to reinforce the lecture material. For example, a typical class period may include some lecture, with an activity or iClicker question/group discussion every 10-15 min.

Formative assessment took place at four time-points (i.e., three mid-terms and a final exam). These exams were open-note, meaning students could use any of their notes – written or printed - on any of the exams. To make sure students weren't using the internet or consulting with any outside help (e.g., group members, friends, family members), we enabled LockDown browser with Respondus monitor for each of the exams. Immediately after a student finished their exam, they could access their score. Each cumulative exam consisted of approximately 50 multiple-choice questions to be completed in two hours' time. Most of the questions were application-based, asking students to interpret phylogenetic trees. Each student's final grade consisted of the two best scores of the three mid-term exams, their final exam score, and participation points gained by answering in-class iClicker questions that mirrored application-based exam questions. We provide final grade information for transparency but note that we only used student exam scores for exam 1, 2, and 3, as individual measures of performance in this study.

Data Collection

After students completed each of their three mid-term exams, we encouraged them to take a voluntary post-exam Qualtrics survey. The survey instrument opened with an information letter detailing the purpose of the study to participants. Then, the survey prompted students to either consent or not consent before moving on to several survey questions. The remainder of the survey collected responses to four 5-point Likert scale questions and one open-response question (Table 3.1). Additionally, at the end of the semester, we downloaded student exam scores to use as a performance measure, so we could compare student study habits for each exam to their performance on that exam.

Table 3.1 The four 5-point Likert scale survey questions and answer choices used in this study.

Likert Scale Survey Items:	**Response Type**
1. Since I had the option to use notes on this exam, my score...	5-Point Likert Scale: (Greatly Decreased (1) to Greatly Increased (5))
2. Since I had the option to use notes on this exam, my anxiety...	
3. Since I was allowed to take this exam using my notes, I think the amount I studied...	
4. Since students in our class were allowed to take this exam using our notes, I think the amount of time other students studied...	
Open Response Survey Items:	
1. How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?	Short Response

Data Analysis

We downloaded survey responses one week after the end of the semester. First, we removed all student data from students who did not consent to be a part of the study. Given that we used many different analytical methods to answer our research questions, we detail each part of the analysis of the Likert scale responses and the open-ended responses in two sections, starting with "Likert Scale Response Analysis". We conducted all analyses and produced all plots for this research in R version 4.0.3 (R Core Team, 2020). We edited plots in BioRender.com.

Likert Scale Response Analysis – Student Perceptions

Student Perceptions: Descriptive Analysis. We calculated response rates for the four Likert scale questions and the open-ended survey question - *How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?* - by dividing the total number of consenting responses by the total number of students enrolled in the course. Next, we assessed Likert scale student responses using a Likert package in R (Bryer, Speerschneider, & Bryer, 2016), calculating the proportion of students reporting a range of responses from "greatly decreased" to "greatly increased" to each of the prompts. We visualized these findings using the plot function.

Student Perceptions: Linear Mixed-Effect Model Analysis. We used the nlme package (Pinheiro et al., 2021) to create linear mixed-effects models, examining the impact of student perceptions of open-note exams on student performance. We first created a model including all four student perceptions measured as our fixed effects (i.e., how the ability to use notes on their exams would impact their exam performance, anxiety levels, the amount they studied, and the amount they thought their peers studied), treating the Likert scale responses as categorical variables. Then, because individual students were sampled longitudinally over the study period, we used a repeated measures design to account for longitudinal measures from a single student by including Student ID as a random effect. After running our first model, we removed one fixed-effect at a time, choosing to remove the fixed-effect with the least significant p-value for each subsequent model. We selected the best fit model using Akaike's Information Criterion (AIC; Akaike, Petrov, & Csaki, 1973), selecting the model with the lowest AIC. In the results section, we included only the relationships that returned significant results, based on a p-value less than 0.05 and confidence intervals that exclude zero.

Next, we created a separate model for each of the four student perceptions measured (i.e., how the ability to use notes on their exams would impact their exam performance, anxiety levels, the amount they studied, and the amount they thought their peers studied), so we could conduct Tukey based post-hoc analysis functions to obtain pairwise significance between exam scores using the emmeans package (Lenth et al., 2018). Then, to visualize all statistically significant relationships, we created customized plots (i.e., combined violin, scatter, and boxplots) in the ggplot2 package (Wickham, 2016).

Open-Ended Response Analysis – Student Study Habits

Student Study Habits: Descriptive Analysis. First, we calculated response rate for the open-ended survey question - *How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?* - by dividing the total number of consenting responses by the total number of students enrolled in the course.

Next, two of the authors (AEB & EPD) individually reviewed all the students' responses to the open-ended question and generated codes using inductive coding (Creswell, 2017). We also took detailed analytic notes at that time (Birks & Mills, 2015). We convened to compare codes and develop one unified coding rubric. Using the unified rubric, AEB and EPD coded a set of 40 responses individually. We met together to compare codes and revise the rubric. We used constant comparison methods to ensure quotes within a category were not too different from each other to warrant the creation of a new theme (Glesne & Peshkin, 1992). This process was repeated until we were confident with the rubric, resulting in the following codes: (a) prepared notes, (b) studied less, (c) understanding, (d) studied the same, (e) less anxious, (f) did not study, (g) external resources, (h) studied more, and (i) no notes (Figure 3.1).

Codes	Code Explanation	Student Excerpt Example
Prepared notes	Student responses mention developing their notes as an exam-taking aid or focusing more energy on their notes.	"Went over my notes and made sure they organized so I would know where to find the correct answers."
Studied Less	The student mentioned they spent less time studying or began studying at a later time point.	"I studied a little less knowing I could look back through my notes to assure an answer."
Understanding	The student focused more on understanding/concepts rather than memorization or details. Alternatively, they relied on notes more than memory.	"I focused more on the overall idea and understanding rather than cramming memorization."
Studied the same	The student studied the same amount or the same way for the open-note exam as they usually do for closed-note exams.	"I studied the same amount."
Less anxious	Less anxiety for this open-note exam as compared to a closed-note exam.	"I was more relaxed."
Did not study	The student admitted to not studying for the open-note exam.	"I did not study really at all for this test, I just took notes during class and used those. If I was going to take it close note, then I would have had to taken the time to study."
External resources	The student mentioned they used external resources such as supplementary instruction session or study guides to prepare for the open-note exam.	"I studied the study guide given by my SI leader which was only available to those students who went to the SI session."
Studied more	The student mentioned they studied more for this open-note exam as compared to a closed-note exam.	"I studied for this more because I knew that it wasn't just about memorizing facts, instead of stressing over remembering every little detail, I spent more time learning and understanding the concepts. Which in turn helped me learn things better."
No notes	Student response stated they did not use notes for this open-note exam.	"I did not study, nor did I use my notes."

Figure 3.1 Emergent codes in student open responses to the open-ended question "*How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?"*

Once the final rubric was established, AEB and EPD individually used the rubric to code the entire data set. Each code is mutually exclusive, meaning an excerpt of text could only be coded as one code. However, students' full responses could include multiple themes. We reached an initial 77.7% agreement, and then met to code to consensus. After we reached consensus, we calculated percentages for each code by dividing the total number of responses assigned for each code by the total number of student responses. After calculating these percentages, we visualized the frequency of each code by constructing bar plots in the ggplot2 package (Wickham, 2016), creating one plot for the codes overall (the average frequency of the code over all three surveys) and one plot for the codes as broken down by each survey timepoint (i.e., exam 1, exam 2, and exam 3). For the overall plot, we included frequencies for all codes, however, for the plot broken down by survey timepoint, we only included information for the codes that demonstrated visual differences between timepoints.

Student Study Habits: Linear Mixed-Effect Model Analysis. To examine the impact of student study habits for open-note exams on student performance, we first reformatted thematic codes as binary data. This means if a student mentioned a particular code in their open-ended response, then they would receive a 1 (yes) for that code. However, if they did not mention a particular code in their open-ended response, then they would receive a 0 (no) for that code. For example, if a student mentioned the code “focusing on understanding” then they were assigned a 1 for that variable. We assigned these values for each qualitative code in the dataset.

After we reformatted the qualitative data into binary data, we created a linear mixed-effects model using the nlme package (Pinheiro et al., 2021) that included all nine student study habit codes as our fixed-effects (i.e., prepared notes, studied less, understanding, studied the same, less anxious, did not study, external resources, studied more, and no notes; Figure 3.1). Then, because individual students were sampled longitudinally over the study period, we used a repeated measures design to account for repeated sampling from a single student during longitudinal measures by including Student ID as a random effect. After running our first model, we removed one fixed-effect at a time, choosing to remove the fixed-effect with the least significant p-value for each subsequent model. We selected the best fit model using AIC (Akaike, Petrov, & Csaki, 1973). In the results section, we only included the study habits that returned significant results. We determined statistical significance and graphically depicted the data as previously described.

RESULTS

Likert Scale Response Results – Student Perceptions

Student Perceptions: Descriptive Results. For the four Likert scale questions (Table 3.1), we received 445 complete responses (i.e., 157 responses after exam 1 [157/218; 72.0% response rate], 147 responses after exam 2 [147/218; 67.43% response rate], and 105 responses after exam 3 [105/218; 48.17% response rate]). After calculating the response rates, we calculated the proportion of students reporting a range of responses from “greatly decreased” to “greatly increased" to each of the prompts. Results demonstrate that students generally equate open-note exams with increased exam scores, decreased anxiety, and decreased studying (Figure 3.2A).

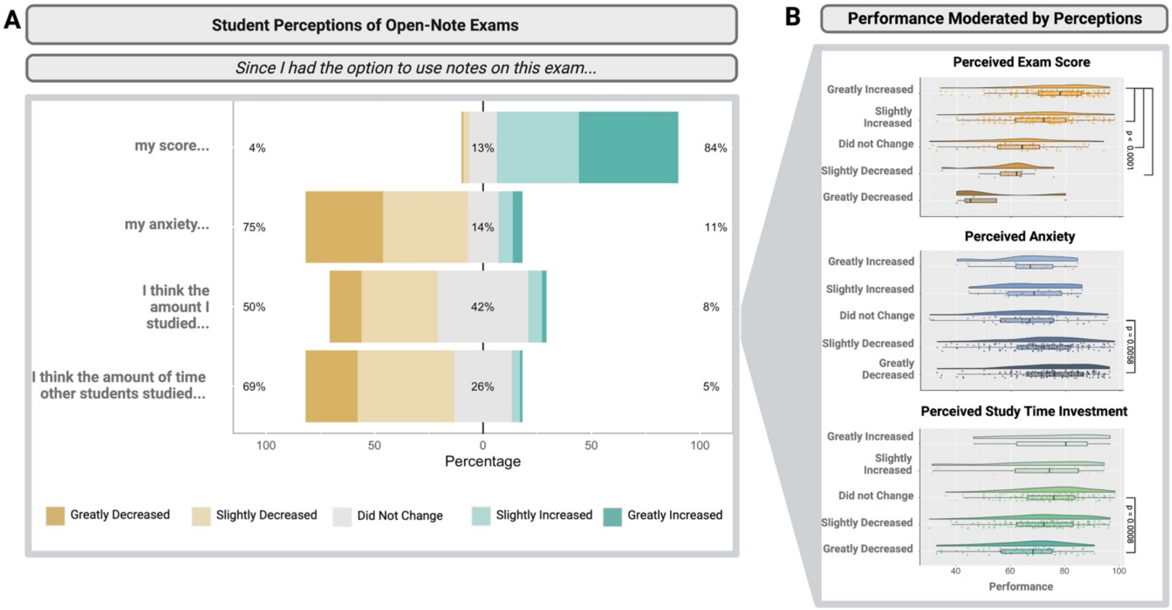


Figure 3.2 *Student Perceptions of Open-Note Exams*. Students were asked a series of 5-point Likert scale questions. (A) Students report perceptions of increased exam performance, decreased anxiety, decreased personal study time investment, and decreased peer study time investment on open-note exams as compared to closed-note exams. (B) Only students reporting that the ability to use notes on their exam "greatly increased" their exam score, "greatly decreased" their anxiety, or "greatly decreased" their study investment had significant impacts on performance.

Student Perceptions: Linear Mixed-Effect Model Results. Next, we examined the relationship between student perceptions of open-note exams and exam performance using linear models. The best fit model for examining the relationship between student perceptions of open-note exams and student performance on those exams included all four student perceptions (i.e., how the ability to use notes on their exams would impact their exam performance, anxiety levels, the amount they studied, and the amount they thought their peers studied) as fixed effects and

Student ID as a random effect. Using this model, three of the Likert scale perceptions displayed clear relationships with student performance: perceived exam score ($F_{(1,202)} = 13.78$, $p < 0.0001$); anxiety ($F_{(1,202)} = 4.62$, $p = 0.0005$); and study time ($F_{(1,202)} = 4.63$, $p = 0.0005$; Figure 3.2B). We then ran a model for each individual perception as a fixed effect with Student ID as a random effect, so we could use these models to conduct Tukey based post-hoc analysis functions. Ultimately, this allowed us to find any differences between Likert scale categories for each individual perception.

We found (1) that students who reported the ability to use notes on their exam "greatly increased" their exam score performed at a significantly higher level than those who claimed their score "slightly increased (7.60 ± 1.37 (SE), $p < 0.0001$), "did not change" (12.79 ± 1.95 (SE), $p < 0.0001$), or "slightly decreased" (18.25 ± 3.55 (SE), $p < 0.0001$); (2) students who reported "greatly decreased" anxiety performed at a significantly higher level than students reporting no change (7.45 ± 2.15 (SE), $p = 0.0058$); and (3) students who reported "greatly decreased" study time investment performed at a significantly lower level than those who reported no change (-8.81 ± 2.15 (SE), $p = 0.0008$). Other incremental comparisons had no significant effect on student performance (Figure 3.2B). We did not observe differences between student responses to the Likert scale questions over time ($p > 0.30$ in each case).

Open-Ended Response Results – Student Study Habits

Student Study Habits: Descriptive Results. For the open-ended survey question - *How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?* - we received 347 total responses (i.e., 140 responses after exam 1 [140/218; 64.22% response rate], 135 responses after exam 2 [135/218; 61.93% response rate], and 72 responses after exam 3 [72/218; 33.03% response rate]). However, we excluded 65 responses (18.68%) from our analysis because they did not fit our codes due to difficulty of interpretation or ambiguity. This left us with 107 coded responses for exam 1, 114 coded responses for exam 2, and 61 coded responses for exam 3.

We identified three important coded responses to the question "*How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?"* Over the three exams, students commonly reported they "prepared notes" for open-note exams. That is, roughly 40% of students mentioned developing their notes as an exam-taking aid or focusing more energy on their notes for these open-note exams than they would have for a closed-note exam (Figure 3.3A). Another common study strategy for open-note exams was to focus on "understanding" the material rather than memorization (21.91%; Figure 3.3A). Lastly, students mentioned accessing "external resources" for studying (2.12%; Figure 3.3A). In our study, we use the term external resources as a catchall term for attending supplemental instruction sessions and using study guides.

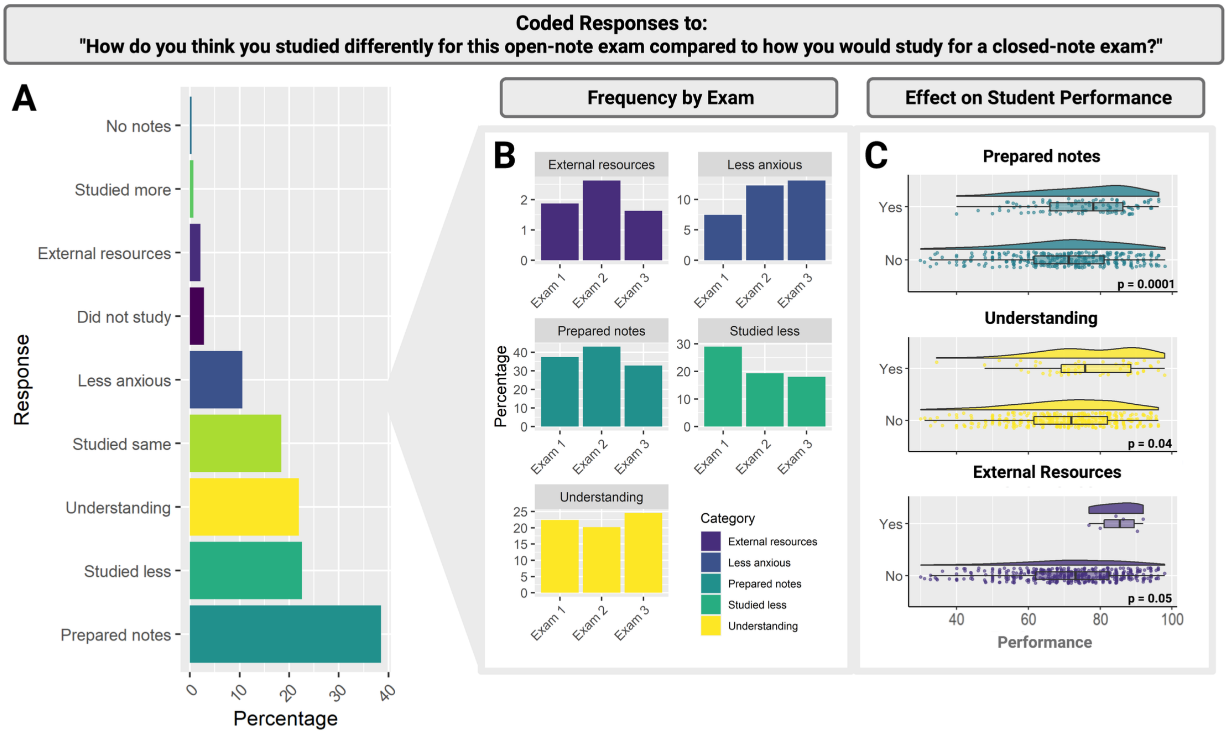


Figure 3.3 *Adapting Study Habits to Open-Note Format*. Students were asked to respond to the following prompt: "*How do you think you studied differently for this open-note exam compared to how you would study for a closed-note exam?"* (A) Student responses are ordered by frequency of student reporting. (B) Longitudinal plotting of student responses indicate student approaches to open-note exams may vary over time in terms of students' use of "External Resources", feeling "Less Anxious", having "Prepared Notes," having "Studied Less" and focus on "Understanding". (C) Student performance increased in response to study habit adaptations that include having "Prepared Notes," an increased focus on "Understanding," and the use of "External Resources."

When we analyzed student responses longitudinally by exam (i.e., exam 1, 2, or 3), only the following five codes varied by exam: (1) external resources, (2) less anxious, (3) prepared notes, (4) studied less, and (5) understanding (Figure 3.3B). While many of these relationships changed over time, the only general trend we found was that, as time went on, students were less likely to report "studying less", which may indicate adaptation to the realities of needing to study for the open-note exam format. Although it should be noted that when students were asked to report their study investment through Likert scale options, their responses did not vary significantly over time. So, while students were less likely to report studying less through short answer responses, we did not identify any variation over time based on quantitative Likert data. Finally, while students often reported “prepared notes” as an adaptation to open-note exams, the frequency of this response decreased for the third exam (Figure 3.3B).

Student Study Habits: Linear Mixed-Effect Model Results. The best fit model for explaining the relationship between student study habits for open-note exams and student performance on those exams included all nine study habit codes (i.e., prepared notes, studied less, understanding, studied the same, less anxious, did not study, external resources, studied more, and no notes) as fixed effects and Student ID as a random effect. Students who focused on preparing notes for open-note exams outperformed students who did not approach studying this way ($F_{(1,229)}$ =14.93, p = 0.0001; Figure 3.3C). Students who focused on understanding to prepare for open-note exams outperformed students who did not approach studying this way ($F_{(1,229)}$ =4.34, p = 0.04; Figure 3.3C). Lastly, students accessing “external resources” for studying outperformed those who did not report this ($F_{(1,229)}$ = 3.85, p = 0.05; Figure 3.3C).

DISCUSSION

We assessed how students perceived open-note exams compared to closed-note exams, whether and how their study habits changed compared to closed-note exams, and how each of these factors impacted student performance. In the following section, we place our findings in the context of current research.

Student Perceptions of Open-Note Exams

Student perceptions of their performance on open-note exams may be consistent with previous literature. For example, Jensen and Moore (2009) showed open-note exams contributed to an overinflated confidence in performance, and this finding may align with our finding that an overwhelming majority of students in our study thought the ability to use their notes would increase their exam score. However, those students who thought the ability to use notes on their exams would “greatly increase” their exam performance significantly outperformed students who thought that same ability would either only “slightly increase,” “not change,” or “slightly decrease” their performance. This may demonstrate students who really thought they would benefit from the exams really did benefit in terms of better performance, potentially demonstrating that those students did not have an overinflated confidence in performance. However, this warrants further investigation, especially because students were able to see their actual score on the exam prior to taking the survey asking about their perceived performance. Additionally, students reported decreased test anxiety levels, which also aligns with previous literature (Block, 2011; Broyles, Cyr, & Korsen, 2005; Gharib, Phillips & Mathew, 2012; Williams & Wong, 2007).

Our finding that students perceived they studied less for their open-note exams, as compared to closed-note exams, is rarely discussed in literature, and could be interpreted in several ways. The most straight-forward interpretation is that students studied the same way for open-note exams as they would for closed-note exams, but for less time. However, students likely studied differently for open-note exams, especially those that test at higher Blooms’ taxonomy levels (e.g., application, analysis, and evaluation; Bloom et al., 1956), and this may have contributed to their perceptions of time spent preparing. For example, students commonly prepared organized notes that were easy to navigate during the exam, relying on those notes as a life raft during the exam. Students who reported studying less for open-note exams may hold a different interpretation of studying, discounting the preparation of notes as a form of studying. Open-note exams alternatively may promote better study habits (Phillips, 2006) given that preparation of notes can be an active method of studying, improving student performance (Rummer et al., 2017; Chen, 2021). Active methods of studying are more effective than passively memorizing facts by re-reading the textbook or study materials, which is time consuming and does not lead to meaningful learning outcomes (Rummer et al., 2017), particularly those outlined in the Vision and Change recommendations (American Association for the Advancement of Science, 2009). While our results cannot explain why students reported studying less, our future work will explore these interpretations of exam preparation.

Three of the Likert scale findings showed relationships with performance outcomes: first, students who reported “greatly increased” perceived exam scores performed at a significantly higher level than those who claimed their score “slightly increased,” “did not change,” or “slightly decreased;” second, students who reported “greatly decreased” anxiety performed at a significantly higher level than students reporting no change; and third, students who reported the amount of time they invested into studying “greatly decreased” performed at a significantly lower level than those who reported no change. Positive performance outcomes were only associated with students who selected the maximum Likert scale responses (i.e., greatly decreased anxiety and study time investment or greatly increased exam score). For example, greatly decreased anxiety positively impacted student performance, but there was no impact of increased or slightly decreased anxiety. Similarly, greatly decreased study time investment negatively impacted student performance, but there was no impact of increased or slightly decreased study time investment on performance. To conclude that student responses to a Likert scale reflect incremental differences in student affect might not fully capture the complex relationships between affect, behavior, and cognition. For example, previous work found a moderate amount of anxiety heightened focus and alertness during assessments, but too much anxiety impaired performance by diverting cognitive resources away from the exam (Wang et al., 2015). Similarly, we may only see performance effects for students reporting more 'extreme' Likert responses, but not among those with milder reports.

Our finding that student perceptions did not significantly shift as experience with open-note exams increased shows that students are not likely to adapt their perceptions of the amount they need to study for an open-note exam as compared to a closed-note exam. Our finding that students' reported anxiety did not change over time suggests that open-note exams can keep student's perceived test anxiety lower than closed-note exams, even after repeated exposure.

Student Study Habits for Open-Note Exams

We found that students who focused on note preparation outperformed their peers, aligning with previous research showing that learning is enhanced by going through the motions of organizing and preparing notes (Erbe, 2007). Note preparation serves as an active and effective form of exam preparation. These findings support previous recommendations for instructors to model active study strategies during class, such as developing structured and accessible notes (Walck-Shannon, Rowell, & Frey, 2021). Our results highlight the importance of modelling the study habits that are effective for this exam type.

Students also focused on "understanding" the material rather than memorization for open-note exams. Students may have written this because they could reference their notes for detail, so they did not need to memorize that information. Rather, they might focus on higher-order concepts, applying the course content to different scenarios. However, students may have been responding to the expectations of the assessment, which, in this research, were different because the exams were open-note. Specifically, the instructors shifted away from a ‘fill in the blank’ style question that students could easily find in their notes. In other words, both students and the instructors may make changes with the knowledge that the exams are open-note: the students in their studying behaviors, and the instructors in how they constructed exam questions. Future research can explore how instructors write open-note exam questions, and how they differ or are like closed-note exams.

Students’ focus on understanding broad concepts rather than memorization supports previous reports showing students prepare for open-note exams by gathering and critically analyzing material from multiple sources as opposed to focusing on storing information for quick retrieval in preparation for a closed-note exam (Theophilides & Koutselini, 2000). Additionally, it lines up with literature showing students expect open-note exams will emphasize understanding and analysis (Eilertsen & Valdermo, 2000). We also found that students who reported focusing on “understanding” to prepare for open-note exams outperformed students who did not report studying this way. These findings, although exploratory, are perhaps the most convincing pieces of evidence we have that open-note exams have the potential to move the goal of undergraduate biology assessments away from memorization and toward deeper understanding and application.

Lastly, we found that students who used “external resources,” such as attending supplemental instruction sessions and using study guides, outperformed those who did not report preparing this way. Previous research shows supplemental instruction, in which a senior student facilitates learning for undergraduate peers in a challenging class, leads to higher grades, lower failure and withdrawal rates, and higher retention and graduation rates (Dawson et al., 2014; Rath et al., 2007). The supplemental instructor in this study, in collaboration with the instructors, developed study guides that outlined the important topics students were expected to know for the exam and included practice questions for students. The supplemental instruction sessions consisted of careful review of the study guide and associated materials, as well as a question-and-answer session.

Our analysis of student responses longitudinally by exam (i.e., exam 1, 2, or 3) showed students reported preparing notes less for the third exam (Figure 3.3B). We hypothesize this decrease could be due to either (1) time demands on the students, given the third mid-term occurred two weeks before final examination week, and students enrolled in our class are often enrolled in many other challenging pre-med classes, or (2) as part of the design of our class, we allowed students to drop one of the mid-term exams, so if they were already doing well in the class, then they may have used the third exam as their drop exam and therefore not studied for it at all, much less with note preparation in mind. These hypotheses, as well as any alternatives, should be tested to delve into the reasoning behind changing study strategies for open-note examinations.

Implications for Practice

We offer several recommendations for undergraduate biology instructors who are considering incorporating open-note examinations into their classrooms. First, we encourage post-secondary biology instructors to ask themselves one key question: what do I want students to gain from their learning and exam experiences? If the objective is for students to focus on understanding rather than memorizing course materials, writing assessment questions that reflect this objective will encourage students to think deeply about the material and reward those who do. Students may benefit from their instructor explaining how their open-note exams are designed to encourage them to specifically focus on understanding, along with the study habits that have previously been shown to increase student performance and learning (e.g., preparing notes, focusing on understanding, and using external resources such as supplementary instruction). Second, we encourage biology education researchers to evaluate how open-note exams are commonly designed and implemented in post-secondary biology classrooms, paying special attention to the Bloom's taxonomy level of each exam question, whether the instructor explains the difference between an open-note and a closed-note exam, and whether the instructor models evidence-based study habits for their students. This will advance our knowledge on the most effective forms of open-note exams, and address why there are mixed-findings in the literature concerning their efficacy.

LIMITATIONS

There are several limitations to the current study. First, we conducted this work amid a global pandemic, which disrupted the lives of students and impacted how they experienced their courses. Students were required to enroll in the remote, online course because of institutional health and safety concerns rather than preference. This means that the student study population, who were taking online, open-note exams, may not represent students who would normally opt for this type of course, potentially impacting our results. Relatedly, the generalizability of these findings is limited because the study took place at a selective, primarily white southeastern institution. Our work in the future will expand beyond this setting, one that is commonly overrepresented in research (Thompson et al., 2020), and address questions related to open-note exams across community colleges, minority-serving institutions, primarily undergraduate institutions, and other student populations. Another limitation includes the fact that the survey we used in our analyses represented single-item survey responses rather than survey constructs. Survey constructs, which are composed of several single-item responses, are preferred because they can be tested for validity by analyzing whether participants answered similarly. Finally, it is important to note that this work examines student perceptions of and study habits for open-note exams that consisted of multiple-choice questions. We acknowledge the importance of work examining perceptions of and study habits for open-note exams consisting of different question-types (e.g., short answer, essay, and design questions), and we note it may produce different results.

CONCLUSIONS

The perceptions students have, and the approaches students take to prepare for an exam can affect performance and (potentially) learning, necessitating the investigation of student exam perceptions and exam preparation relevant to the exam type. However, investigations such as these have yet to be conducted with novel types of exams such as open-note exams. Our research on one such novel exam, open-note exams, indicates that (1) students perceived they would perform better on open-note exams while studying less and experiencing less anxiety, (2) study habits including focusing on note-preparation, deep understanding, and application may lead to higher performance in biology classes that use open-note exams, (3) students often practice unsuccessful study habits, so they may benefit from instructor guidance on effective preparation techniques, and (4) in order to effectively assess open-note exams, future research study designs must take into account the appropriateness of the exam design. We emphasize that low-level Bloom's taxonomy, closed-note, multiple choice question exams do not effectively address Vision and Change needs for student success. However, when properly designed, open-note exams can promote in-depth, higher-level thinking, as well as effective study habits.

ACKNOWLEDGEMENTS

We thank Robin Costello, Sharday Ewell, Enya Granados, Taylor Gussler, Todd Lamb, and Ariel Steele of the Ballen lab for their helpful guidance in editing this manuscript. Additionally, we are grateful to the biology students for their participation in this study.

REFERENCES

Agarwal PK, Karpicke JD, Kang SHK, Roediger HL, McDermott KB (2008). Examining the testing effect with open- and closed-book tests. Appl Cogn Psychol 22, 861–876.

Agarwal, P. K., & Roediger, H.L. III. (2011). Expectancy of an open-book test decreases performance on delayed closed-book test. Memory, 19(8), 836-852.

Akaike, H., Petrov, B. N., & Csaki, F. (1973). Second international symposium on information theory.

Alpert, W. T., Couch, K. A., & Harmon, O. R. (2016). A randomized assessment of online learning. American Economic Review, 106(5), 378-82.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). How learning works: Seven research-based principles for smart teaching. John Wiley & Sons

American Association for the Advancement of Science. (2009). Vision and change: A call to action. Washington, DC. Retrieved January 29, 2020, from https://live-visionandchange.pantheonsite.io/wp-content/uploads/2011/03/VC-Brochure-V6-3.pdf

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. PLoS One, 12(10), e0186419.

Bettinger EP, Fox L, Loeb S, Taylor ES. 2017. Virtual classrooms: how online college courses affect student success. American Economic Rev 107:2855–2875.

Birks, M., & Mills, J. (2015). Grounded theory: A practical guide. Sage.

Blaisman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behavior. Memory, 25(6), 784-792.

Block RM (2011). A discussion of the effect of open-book and closed- book exams on student achievement in an introductory statistics class. PRIMUS 22, 228–238.

Bloom, B. S., Englehard, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Committee of College and University Examiners. Taxonomy of educational objectives: the classification of educational goals. New York, 16, 207.

Broyles, I.L., Cyr, P.R. and Korsen, N. 2005. Open book exams: Assessment of academic learning in clerkships. Medical Teacher, 27(5): 456–462.

Cassady, J., & Johnson, R. (2002). Cognitive test anxiety and academic performance. Contemporary Educational Psychology, 27(2), 270-295. http://dx.doi.org/10.1006/ceps.2001.1094

Chapell, M., Blanding, B., Silverstein, M., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. Journal of Educational Psychology, 97(2), 268-274. http://dx.doi.org/10.1037/0022-0663.97.2.268

Chen, P. H. (2021). In-class and after-class lecture note-taking strategies. Active Learning in Higher Education, 22(3), 245-260.

Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.

Dawson, P., van der Meer, J., Skalicky, J., & Cowley, K. (2014). On the effectiveness of supplemental instruction: A systematic review of supplemental instruction and peer-assisted study sessions literature between 2001 and 2010. Review of educational research, 84(4), 609-639.

Eilertsen, T. V., & Valdermo, O. (2000). Open-Book Assessment: A Contribution to Improved Learning?. Studies in Educational Evaluation, 26(2), 91-103.

England, B. J., Brigati, J. R., & Schussler, E. E. (2017). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. PloS one, 12(8), e0182506.

England, B. J., Brigati, J. R., Schussler, E. E., & Chen, M. M. (2019). Student anxiety and perception of difficulty impact performance and persistence in introductory biology courses. CBE—Life Sciences Education, 18(2), ar21.

Ewell, S. N., Josefson, C. C., & Ballen, C. J. (2022). Why Did Students Report Lower Test Anxiety during the COVID-19 Pandemic?. Journal of Microbiology & Biology Education, e00282-21.

Feller, M. (1994). Open-Book Testing and Education for the Future. Studies in Educational Evaluation, 20(2), 235-38.

Gharib A, Phillips W, Mathew N (2012). Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety. Psychol Res 2, 469–478.

Glesne, C. & Peshkin, A. (1992). Becoming qualitative researchers. New York: Longman Publishing Group.

Green, S. G., Ferrante, C. J., & Heppard, K. A. (2016). Using Open-Book Exams to Enhance Student Learning, Performance, and Motivation. Journal of Effective Teaching, 16(1), 19-35.

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., ... & Dilig, R. (2020). The Condition of Education 2020. NCES 2020-144. National Center for Education Statistics.

Ioannidou, M. K. (1997). Testing and life-long learning: Open-book and closed-book examination in a university course. Studies in Educational Evaluation, 23(2), 131-139.

Jensen PA, Moore R. 2009. Students' perceptions of their grades throughout an introductory biology class: effect of open-book exams. J Coll Sci Teach 38:58–61.

King, C. G., Guyette Jr, R. W., & Piotrowski, C. (2009). Online exams and cheating: An empirical analysis of business students' views. Journal of Educators Online, 6(1), n1.

Krasne S, Wimmers P, Relan A, Drake T (2006). Differential effects of two types of formative assessment in predicting perfor-mance of first-year medical students. Adv Health Sci Educ 11, 155–171.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. R package version, 1(1), 3.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE—Life Sciences Education, 9(4), 435-440.

Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. CBE—Life Sciences Education, 12(2), 239-249.

Moore R, Jensen PA. 2007. Do open-book exams impede long-term learning in introductory biology courses? J Coll Sci Teach 36:46–49.

Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020). Minimize online cheating for online assessments during COVID-19 pandemic. Journal of Chemical Education. https://doi.org/10.1021/acs. jchemed.0c00790

Phillips G. 2006. Using open-book exams to strengthen the study skills of community-college biology students. J Adolesc Adult Literacy 49:574–582.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152, https://CRAN.R-project.org/package=nlme.

R Core Team (2020). R: A language and environment for statistical## computing. R Foundation for Statistical Computing, Vienna, Austria. ## URL https://www.R-project.org/.

Rath, K. A., Peterfreund, A. R., Xenos, S. P., Bayliss, F., & Carnal, N. (2007). Supplemental instruction in introductory biology I: Enhancing the performance and retention of underrepresented minority students. CBE—Life Sciences Education, 6(3), 203-216.

Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? Journal of Experimental Psychology: Applied, 23(3), 293–300. https://doi.org/10.1037/xap0000134

Salehi, S., Cotner, S., Azarin, S. M., Carlson, E. E., Driessen, M., Ferry, V. E., ... & Ballen, C. J. (2019). Gender performance gaps across different assessment methods and the underlying mechanisms: The case of incoming preparation and test anxiety. In Frontiers in Education (p. 107). Frontiers.

Sato BK, He W, Warschauer M, Kadandale P. 2015. The grass isn’t always greener: perceptions of and performance on open-note exams. *LSE* 14:ar11.

Seymour E, Hewitt NM (1997). Talking about Leaving: Why Undergraduates Leave the Sciences, Boulder, CO: Westview.

Seymour, E., Hunter, A. B., Harper, R. P., & Holland, D. G. (2019). Talking about leaving revisited. Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education.

Szafranski, D. D., Barrera, T. L., & Norton, P. J. (2012). Test anxiety inventory: 30 years later. Anxiety, Stress, and Coping, 25(6), 667-677. http://dx.doi.org/10.1080/10615806.2012.663490

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open-book exam: A comparative analysis. Educational Research and Evaluation, 6(4), 379-393.

Thompson, S. K., Hebert, S., Berk, S., Brunelli, R., Creech, C., Drake, A. G., Fagbodun, S., Garcia-Ojeda, M. E., Hall, C., Harshman, J., Lamb, T., Robnett, R., Shuster, M., Cotern, S., & Ballen, C. J. (2020). A call for data-driven networks to address equity in the context of undergraduate biology. CBE—Life Sciences Education, 19(4), mr2.

Walck-Shannon, E. M., Rowell, S. F., & Frey, R. F. (2021). To What Extent Do Study Habits Relate to Performance?. CBE—Life Sciences Education, 20(1), ar6.

Wang, Z., Lukowski, S. L., Hart, S. A., Lyons, I. M., Thompson, L. A., Kovas, Y., ... & Petrill, S. A. (2015). Is math anxiety always bad for math learning? The role of math motivation. Psychological science, 26(12), 1863-1876.

Watson, G. R., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses?.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Williams JB, Wong A (2007). Closed book, invigilated exams versus open book, open web exams: An empirical analysis. In: ICT: Pro- viding Choices for Learners and Learning.

Proceedings Australasian Society for Computers in Learning in Tertiary Education Singapore 2007, ed. RJ Atkinson, C McBeath, SKA Soong, and C Cheers, Centre for Educational Development, Nanyang Technological University, Singapore, 2–5 December, 1079–1083. http://www.ascilite.org.au/ conferences/singapore07/procs/.

Wood WB (2009). Innovations in teaching undergraduate biology and why we need them. Annu Rev Cell Dev Biol 25, 93-112.

Xu, D., & Jaggars, S. (2011). Online and hybrid course enrollment and performance in Washington State community and technical colleges.

Xu D., & Jaggars S. 2014. Performance gaps between online and face-to-face courses: differences across types of students and academic subject areas. The J Higher Education 85:633–659.

Xu, J., Ong, J., Tran, T., Kollar, Y., Wu, A., Vujicic, M., & Hsiao, H. (2021). The impact of study and learning strategies on post-secondary student academic achievement: A mixed-methods systematic review.

Vitasari, P., Wahab, M. N. A., Othman, A., Herawan, T., & Sinnadurai, S. K. (2010). The relationship between study anxiety and academic performance among engineering students. Procedia-Social and Behavioral Sciences, 8, 490-497.

Zeidner, M. (1998). Test anxiety: The state of the art.

Zheng, Alex Y., Janessa K. Lawhorn, Thomas Lumley, Scott Freeman, Application of Bloom's Taxonomy Debunks the``MCAT Myth''. Science,319,414-415. (2008).

Chapter 4: Learning principles of evolution during a crisis: An exploratory analysis of student barriers one week and one month into the COVID-19 pandemic

Emily Driessen*, Abby Beatty, Alexis Stokes, Sara Wood, & Cissy Ballen
Department of Biology, Auburn University, Auburn, AL 36849

*Corresponding Author. Contact @ epd0016@auburn.edu

Running Header: Learning evolution during a crisis.

This manuscript was accepted and published as a research article with Ecology and Evolution in 2020.

ABSTRACT

The Coronavirus disease (COVID-19) outbreak forced an emergency transition to online classes across the world with little warning or instruction for faculty and students. The goal of this research was to document how this response impacted undergraduate students studying the principles of evolution in an introductory organismal biology class over time; specifically, how their study habits for exams differed (1) one week and (2) one month after a university's decision to transition to emergency remote instruction. We asked students about the extent to which COVID-19 impacted their study habits, and we categorized students' responses using open coding. We identified several consistent similarities - as well as dramatic differences - in their responses as the time away from campus increased. The report that follows is a summary of the documented barriers and recommendations based on literature concerning crises and equitable practices.

BACKGROUND

In the last decade, college students have dealt with tragedies or disasters on national, local, and individual levels. These incidences have strong emotional and cognitive impacts on students - even those who are not directly affected (Honos-Webb, Sunwolf, Hart, & Scalise, 2006). Examples include natural disasters such as Hurricane Maria, mass shootings (e.g. at the University of Texas, University of Iowa, and Virginia Tech; Post, 2016), local tragedies such as the murder of a fellow student (Mettler, 2020), as well as individual tragedies such as the sudden death of a family member. In response to the widespread prevalence of such tragedies, previous work documents campus actions that provide emotional support. For example, crisis response teams at some institutions provide support to anticipate or respond to tragedies (Asmussen & Creswell, 1995; Honos-Webb et al., 2006). Other institutions have secured venues specifically for students and personnel to congregate and learn more about an incident and process grief (Hurst, 1999).

Although the response to these disasters and tragedies are well documented, the COVID-19 pandemic and its implications for higher education presented unique challenges and research opportunities to inform future learning transitions. For example, the COVID-19 outbreak resulted in an emergency transition to online classes. Learning in this context presented practical and complex obstacles—including difficulty accessing online course materials due to intermittent internet access, processing the illness of a close family member, and preference for in-person instruction. Teaching in this context also presented practical obstacles for faculty - online instruction requires a different skillset than in-person instruction (Rohland-Heinrich, 2016; Slovick, 2011). These skills take time and guidance to personally develop, neither of which were available for many instructors during the rapid transition. Additionally, a unique feature of the COVID-19 response was the call for students to shelter in place for an indefinite period, cutting off contact to study groups and collaborative learning opportunities in class. Thus, for students enrolled in challenging, large science classes - particularly those that rewarded group work and participation - closure of the institution and self-isolation resulted in a learning environment contradictory to the collaborative classroom settings to which students had grown to expect. Since the COVID-19 pandemic erupted mid-semester, instructors navigated multiple waves of challenges immediately after the closure of institutions through the end of the semester. This led us to investigate the following research question: How did the institutional transition to online instruction, a result of the COVID-19 pandemic, affect student study habits over time?

To consider how COVID-19 impacted students over time, we set out to document challenges students faced as they studied for online exams (1) one week and (2) one month after our institution's decision to transition to remote instruction. Our study took place in the context of a large, organismal biology class with an explicit focus on organismal evolution and classification. This course was taught at a southeast university in the United States (U.S.), so it is important to note the conclusions of this study may not be generalizable outside of this particular region or country. By using a repeated measures design, we captured collective snapshots of our students' condition as they forcibly adapted to remote learning during this global emergency. We hypothesized that student challenges would markedly differ at these two time points, but some challenges would persist throughout the semester. For example, one week after the transition online we hypothesized students would report challenges related to the chaos of moving off campus and adjusting to online content delivery. This is because at that time, they had not grappled with learning content online for an extended time. In contrast, we expected that one month after the transition online, students may report persistent barriers they faced while adapting to the new norm of online coursework. By conducting exploratory studies that document the student experience we contribute to a knowledge base that can inform institutional policy-makers as well as instructors as we prepare for the future.

METHODS

Course background

We collected data in an Organismal Biology course from spring 2020, with a total enrollment of 415 students across two class sections. This class is the second in an introductory biology sequence taught at a research intensive, land-grant university in the southeast region of the United States, with two of the paper's authors as the instructors of record (AB & CJB). The principle objective of the course is for students to develop an understanding of the evolution, classification, structure, and spectacular diversity of living organisms, focusing on plants and animals. This course is designed for those in biology-related majors to prepare them for future coursework. In this course, students were assigned to small groups (ranging from four to six members) at the beginning of the semester and they sat and worked together for every class. It is a three-credit class that includes two 75-minute class sessions each week, and is offered every semester. The course initially met in a large auditorium, with seating for 270 students.

Each class period included lecture and group activities to reinforce the lecture material. For example, a typical class period may include some lecture, with 3-4 iClicker questions/group discussions every 15 minutes, as well as a 20 minute group activity. Students also took a group exam following individual exams, allowing for students to confer with their group members over exam materials.

During the first ten weeks of the semester, students had completed the first unit, including both the lessons and the first exam, and the lesson material for the second exam. The students were then released for spring break with the plan to take the second exam in person upon returning from break. However, unpredictably, the university cancelled in-class meetings, four days before students were to return from spring break, due to the COVID-19 pandemic. For that reason, class instruction became fully online four days after the cancellation of in-class meetings, and instruction was to remain online for the rest of the semester (i.e. 7 weeks). Students then took their second exam one week after the transition online and took their third exam one month following the transition to the online format. Following the transition to online teaching, the instructors made all pre-recorded 'mini-lectures' available, via Canvas, to students for the rest of the semester, and required students complete a 10-question 'quarantine quiz' that corresponded to each class period. Of note, neither instructor had formal experience teaching online, but both had some experience with taking online classes as students.

Data Collection

Immediately after students completed exams, both one week and one month after the transition online, we encouraged students to take a voluntary post-exam survey online via Qualtrics in exchange for a small amount of extra credit (awarded for clicking on the link to the survey). We selected an open-ended survey question that could not be answered with a simple yes or no. Rather, we probed for elaboration concerning changes in student study habits: "to what extent did the Coronavirus disease (COVID-2019) impact your study habits?" Our study offered an unusual opportunity to examine students' responses after quickly transitioning from one type of instruction to another, rather than opting-in to a particular mode of content delivery.

In this repeated measures study, we measured the same variable(s) from the same population at two or more time points (i.e., one week and one month after the online transition).

Data Coding and Analysis

We downloaded survey responses one week following implementation. We found 330 students responded to both survey items (80% of students). For the first dataset, two of the authors (CJB & EPD) created categories through first- and second-cycle analyses (Saldana, 2015) using open and thematic coding, and the following six codes emerged: (1) focus/motivation/confusion, (2) cannot access campus resources, (3) cannot access group members, (4) no change, (5) less time, and (6) more time (Table 4.1). The responses to the survey question of concern (i.e., to what extent did the Coronavirus (COVID-19) impact your study habits?) were assigned as many codes as fit, meaning they could fit more than one code. Less than 5% of responses did not match any of the codes, and they were left uncoded. Two coders coded to consensus, meaning both coders agreed on code assignments for all responses. We calculated percentages for each code by dividing the total number of responses assigned for each code by the total number of student responses.

Table 4.1 Emergent themes in student open responses.

Code	Definition: Students...	Example
Focus/motivation/confusion	...reported a variety of emotions, which in turn impacted productivity/learning	“Having to leave and not being able to have a quiet place to study at home living with a family of six took focus away from studying.”
Cannot access campus resources	...reported that physical study materials or access to the instructor or learning assistants were not available	“I enjoyed going to the library to study. I can no longer do that.”
Less time	...had less time to study and focus on their school work	“My children are also out of school, and I am balancing my time between ensuring that they are doing what they

		need to for classes and doing my own school work.”
No change	...were not impacted by the pandemic	“It did not impact my study habits.”
Cannot access group	...experienced fewer in-person meetings with group members, with whom they were required to sit in class	“I was unable to study with friends and had no way of knowing if my knowledge had holes in it. Collaboration is key to my academic success.”
More time	...had more time to study and focus on their school work	“It actually helped because it gave me more time to chill and less time to travel.”
Routine change	….experienced a disruption in daily habits or structure	“I've become more of a procrastinator without a normal routine.”
Sub-optimal learning	...experienced a less than ideal learning environment where they studied away from campus	“I am not used to being home and studying around my family. My family is very loud and I can never get any peace or quiet. I am not able to go to my usual designated study spots or even leave the house at all.”
Harder to learn online	...were challenged by online learning, compared to in-person instruction	“COVID 19 made studying a little harder because online lectures are distracting for me.”
Easier to learn online	...preferred online learning, compared to in-person instruction	“I found it easier to be able to understand the concepts more when having the lectures recorded and in front of me.”
Illness	...had family members, close friends, or they were ill	“COVID-19 impacted mine because I have been worried about my immunocompromised family member getting sick, and I think I'm already sick from going through the airports. I've been focused on helping my family, and I've had to be the one running the errands.”
Other	Responded in a way that was either vague or didn’t address the impact COVID-2019 had on their study habits.	“It has made me study a lot more.”

In response to the one month survey, we found 296 students responded to both survey items (71% of students). Four of the authors (EPD, AS, SW, & CJB) created an additional 5 codes to account for new emergent themes: (1) routine change, (2) sub-optimal learning environment, (3) more difficult to learn online, (4) illness, and (5) easier to learn online. The same two coders coded to consensus. After the one month data were coded, the two coders reanalyzed the one-week dataset with the additional 5 categories and coded to consensus again. Then, we calculated percentages for each code.

RESULTS & DISCUSSION

Results consisted of one survey question answered at two time points: one week and one month after the university's transition to online-instruction in response to the COVID-19 pandemic. We received a total of 330 survey responses one week after the transition (80% of students) and 296 survey responses one month after the transition (71% of students). We asked "to what extent did the Coronavirus disease (COVID-2019) impact your study habits?" and developed categories that encompassed most student responses (Table 4.1). The three most common student responses one week after transitioning online included focus/motivation/confusion (37.58% of responses), less time to study (17.27%), and not being able to access campus resources (17.27%); the three most common responses one month after transitioning included focus/motivation/confusion (34.12% of responses), a sub-optimal learning environment (28.72%), and it's more difficult to learn online (16.22%) (Table 4.2, Figure 4.1).

Table 4.**2**. Comparison of Categorized responses to the prompt, "To what extent did the Coronavirus disease (COVID-19) impact your study habits?" at two time points: one week after and one month after the university switched to online-only instruction in response to the COVID-19 outbreak. The six most common emergent themes for both time points are highlighted

highlighted in orange (one week data) and blue (one month data). The percentages represent the number of students who mentioned the category as divided by the total number of student respondents.

Category	One week	One month
Focus/Motivation/Confusion	37.58%	34.12%
Cannot Access Campus Resources	17.27%	11.82%
Less Time	17.27%	5.07%
No Change	14.85%	9.78%
Cannot Access Group	11.82%	11.82%
More Time	6.06%	3.04%
Routine Change	1.21%	12.16%
Sub-optimal learning environment	0.91%	28.72%
Harder to learn online	0.30%	16.22%
Easier to learn online	0.30%	3.04%
Illness	0.30%	1.69%

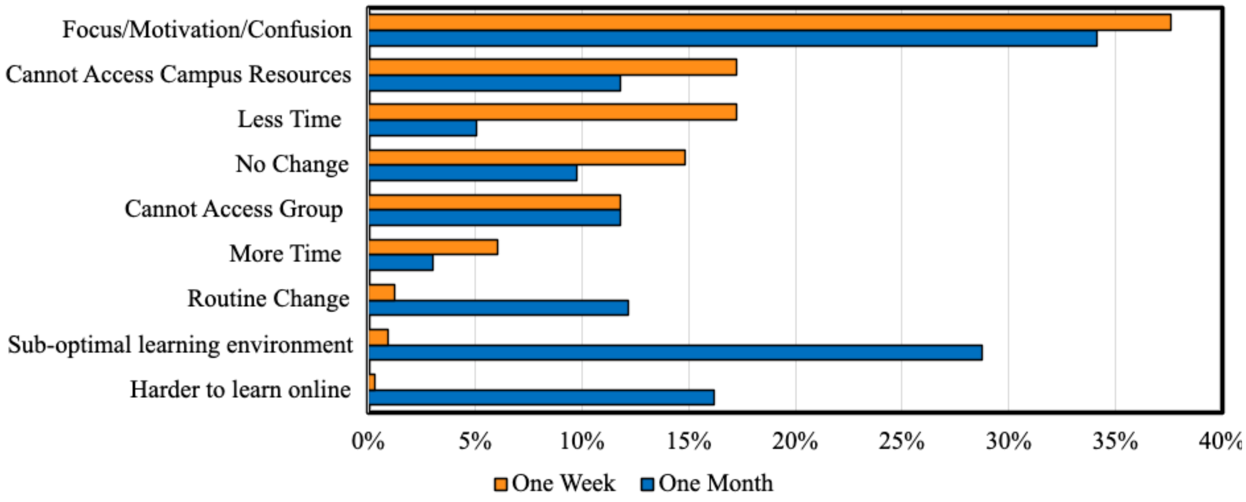


Figure 4.1 Comparison of categorized student responses to the prompt, “To what extent did the Coronavirus disease impact your study habits?” one week (orange) and one month (blue) after the COVID-19 transition online. All responses were assigned as many codes as relevant, meaning they could fit more than one code. Percentages were then calculated for each code by

dividing the total number of responses assigned for each code by the total number of student responses. This figure displays only the six highest percentages for each dataset. Values indicate the percentage of students who reported each category as an impact of the institutional changes.

Several themes differed between student responses one week and one month after a switch to online instruction. For example, one week after the transition to online learning, students emphasized challenges related to the chaos of moving off campus (e.g., “It was harder to study because we had to travel home and pack up things from our dorms” and “[The COVID-19 pandemic] created a sense of panic and caused distraction from studying”)(Table 4.2; Figure 4.1). On the other hand, one month after the transition online, students emphasized (1) issues related directly to online instruction (e.g., “Harder to study because all of the material had to be learned online and by myself”); (2) a sub-optimal learning environment (e.g., “I cannot focus at my house and I have three siblings and they interrupt me during my tests and I cannot focus”); or (3) the lack of structure in their daily routine (e.g., “it broke my class-library-work routine, and it has been more difficult to stay on top of my work without this kind of organized self-discipline”). These differences made sense, because after one week students had just moved off campus and were adjusting to the online learning environment. After one month, the prospects of returning to campus were waning. At this time, students were sheltering in place, and expected to interface with a virtual classroom - something many of them had never experienced in their education.

Though different patterns emerged when we compared student responses across the two time points, some themes dominated throughout the entire study period. One recurring theme from students was the negative impact of emotional factors on their study habits, such as anxiety, a sense of confusion, and lack of motivation (Table 4.2; Figure 4.1). While these were all placed in the same category during coding, the detailed responses put them in a more nuanced context; and differed one week or one month after transitioning online. For example, one week after transitioning online, students reported lacking motivation to study, with one student explaining that they were "less worried about this exam because there are a lot of more important things going on in my life right now." It was common for students who cited a lack of motivation one month after the transition online to also describe distractions at home or the lack of a routine/structure. For example, "I have very little motivation to get things done at home since there is not really any structure." This result highlights the persistent and potentially damaging impact that emotional factors have on student learning and well-being during an extended crisis period.

Our results yield implications of this work that are two-fold: exploratory studies can inform (1) institutional and (2) instructional policies in times of crisis.

First, institutions that seek feedback in the wake of COVID-19 can develop an informed response in the case of future emergencies. For example, students in our sample reported COVID-19 negatively impacted their emotional well-being, which in turn affected their study habits, and this persisted over time. Under normal circumstances - those uncomplicated by a global pandemic - 23.2% and 15.4% of American college students' academic performances were negatively impacted by anxiety and depression, respectively (American College Health Association, 2016). To prepare for the additional emotional struggle experienced during a crisis, some institutions have developed emergency response teams to support or respond (Honos-Webb et al., 2006; Asmussen & Creswell, 1995). At our institution, and presumably others, emotional support opportunities through the student counseling and psychological services closed for physical safety in response to the virus, even after instruction resumed. Thus, students were not able to obtain emotional or psychological support during the height of the crisis when it would have benefited the most students. Our results highlight the importance of this resource.

The second implication of this work is that exploratory studies can inform instructional practices in times of crisis. Based on personal communications with instructors, many experienced an uptick in student emails detailing logistical or emotional struggle while sheltering in place. However, these direct communications come from only a subset of students who are comfortable reaching out to an instructor (Miller & Pearson, 2013). We collected data from most students in the class (i.e., between 70-80% of students participated), decreasing the chance of a biased representation of student input. By formalizing data collection and using qualitative coding, where possible, instructors can better understand the most salient challenges for students and respond by adapting their course materials. In addition to aiding in evidence-based instructional decisions, previous work also shows students respond positively when they perceive instructors care about their wellbeing and exhibit warmth and respect for students (Dewsbury & Brame, 2019). Sharing anonymous student feedback from the survey with the class also has the potential to foster a sense of community (Phelan, 2012).

CONSIDERATIONS AND FUTURE DIRECTIONS

We are aware that instructors everywhere have been faced with countless examples of struggle through communications with students. This study offers insights into student experiences from two sections of one introductory biology course at a large southeastern university. While our survey was imperfect, the simple question about COVID impacts yielded meaningful insights. Similar work should be conducted in different geographic locations and at different institution types, where students might be facing different challenges. For example, the state in which the current research was conducted implemented a delayed stay-at-home order relative to other states, and lifted the order relatively early. In contrast, California was the first state to transition online and ask its residents to shelter in place. We may expect students taking similar coursework to be differently impacted due to these circumstances, and therefore optimal instructor responses and institutional supports might also differ.

CONCLUSIONS

Our study exposed the tremendous strain COVID-19 placed on students and their families. Teaching during this crisis confronted instructors with inequities that are undeniable, with differences in access to technology and levels of financial stability; students reported caring for siblings if they had parents who were essential workers, or caring for their children who could not attend child care. Students themselves took on essential jobs to support themselves or their families. Locally, Alabama, the state in which our institution is located, experienced secondary weather emergencies that left students and their families without shelter. By offering this simple question to students, we gained a glimpse into their lives and observed stark differences based on socioeconomic status. These inequities underscore the importance of implementing equitable institutional policies and teaching practices to ensure that all students can excel, especially in a time of crisis. We hope that this exploratory study can serve as a helpful resource for researchers and instructors who are interested in researching barriers or developing a response for their students.

ACKNOWLEDGEMENTS

We would like to thank Sara Berk, Ísada Claudio-Ford, Todd Lamb, Sara Odom, and Ariel Steele for providing critical feedback on earlier drafts of this study, as well as two anonymous reviewers who provided helpful comments that greatly improved the quality of this work. This research was approved by the Institutional Review Board, Protocol #20-189 EX 2004.

DATA ACCESSIBILITY STATEMENT

Due to ongoing research with this dataset as well as IRB restrictions, data is not publicly available but is available upon request.

COMPETING INTERESTS

The authors of this article have no competing interests that would impact the integrity of this work.

AUTHOR CONTRIBUTIONS

Emily Driessen: Conceptualization (Lead), Data curation (Equal), Formal analysis (Supporting), Methodology (Equal), Writing-original draft (Lead), and Writing-review & editing (Lead)

Abby Beatty: Writing-original draft (Supporting) and Writing-review & editing (Supporting)

Alexis Stokes: Formal analysis (Lead)

Sara Wood: Formal analysis (Lead)

Cissy Ballen: Conceptualization (Lead), Data curation (Equal), Investigation (Supporting), Methodology (Supporting), Writing-original draft (Equal), and Writing-review & editing (Supporting)

REFERENCES

American College Health Association (2016). American College Health Association National College Health Assessment NCHA II, Spring 2016 Reference Group Data Report.

Hanover, MD.Asmussen, K. J., & Creswell, J. W. (1995). Campus response to a student gunman. The Journal of Higher Education, 66(5), 575–591.

Dewsbury, B., & Brame, C. J. (2019). Inclusive teaching. Cbe—life Sciences Education, 18(2), fe2. https://doi.org/10.1187/cbe.19-01-0021

Honos-Webb, L., Sunwolf, Hart, S., & Scalise, J. T. (2006). How to help after national catastrophes: Findings following 9/11. The Humanistic Psychologist, 34(1), 75–97. https://doi.org/10.1207/s15473333thp3401_7

Hurst, J. C. (1999). The Matthew Shepard tragedy: Management of a cri-sis. About Campus, 4(3), 5–11. https://doi.org/10.1177/1086482299 00400303

Mettler, K. (2020). A 14-year-old has been charged with murder in the stabbing of Barnard College student. Retrieved from https://www.washingtonpost.com/nation/2020/02/15/tessa-majors-barnard-college-arrest/

Miller, A., & Pearson, J. (2013). Can I talk to you? The effects of instructor position, nationality, and teaching style on students' perceived will-ingness to communicate and on teacher evaluations. Communication Quarterly, 61(1), 18–34. https://doi.org/10.1080/01463373.2012.719059

Phelan, L. (2012). Interrogating students’ perceptions of their online learning experiences with Brookfield’s critical incident questionnaire. Distance Education, 33(1), 31–44. https://doi.org/10.1080/01587919.2012.667958

Post, C. (2016). The deadliest college campus shootings in U.S. history. Retrieved from https://www.daytondailynews.com/news/crime%2013law/the-deadliest-college-campus-shootings-history/bx9GETTkckx0on6k8QiycN/

Rohland-Heinrich, N. L. (2016). Transitioning from lectern to laptop: Faculty experiences in online instruction. Doctoral dissertation.

Saldana, J. (2015). The Coding Manual for Qualitative Researchers. Newcastle upon Tyne, UK: Sage.

Slovick, E. A. (2011). A descriptive qualitative study of instructors' perceptions in transitioning from a traditional classroom to online instruction. Doctoral dissertation, University of Phoenix.

CONCLUSION

Here, I explored, defined, and tested the impact of varying course elements. Specifically, I (1) defined the term “active learning” and catalogued the 300+ strategies used and researched in the context of biology education instruction, (2) used meta-analysis to show group work has the potential to increase student performance in post-secondary biology courses by approximately one letter grade (3) proved open-note exams lower student perceived exam-anxiety and cause a shift in student study habits, often promoting higher performance in the case of those students who began focusing on note preparation and broad conceptual understanding rather than rote memorization, and (4) documented how the emergency transition to online classes, due to the Coronavirus disease (COVID-19) outbreak, impacted undergraduate student study habits in an introductory organismal biology class over time. The four chapters combined expand our understanding of course elements and their effects on student affect and student performance, both at a small-scale (i.e., in one course at a single institution in the case of Chapters 3 and 4) and a large-scale (i.e., from multiple studies at multiple institutions in the case of Chapters 1 and 2). We found that course elements such as group work and open-note exams can have positive impacts on student affect (i.e., support better study habits and decrease test anxiety in the case of open-note exams) and performance (i.e., increase student grades by an average of one letter grade in the case of group work). On the other hand, course elements such as a sudden change in access to in-person resources, in the case of the emergency transition to remote-online instruction, can negatively impact student study habits. The idea that particular course elements can help students while others can harm them is in line with many other recent studies I have conducted on course elements, including research on novel curriculums (Beatty et al., 2021a; Adams et al., in review; Beatty et al., in revision), differing levels of support during course-based undergraduate research experiences (Beatty et al., 2021b), exams (Tracy et al., 2022), course format (Tracy et al., 2022), course workload (Tracy et al., 2022), and course content (Tracy et al., 2022).

In Chapter 1, we defined active learning empirically through a systematic literature review of three biology education research journals (n = 148 articles) and a survey of a popular biology education research community (i.e., the Society for the Advancement of Biology Education Research; SABER; n = 105 individuals). Our objectives were to increase transparency and reproducibility of teaching practices and research findings in biology education. Findings showed much of the literature concerning active learning never defined the term, but the authors often provided examples of specific active-learning strategies. We categorized the available active-learning definitions and strategies obtained from the articles and survey responses to highlight central themes. Based on data from the BER literature and community, we provide a working definition of active learning and an Active-Learning Strategy Guide that defines 300+ active-learning strategies. These tools can help the community define, elaborate, and provide specificity when using the term active learning to characterize teaching practices.

In Chapter 2, we tested the isolated impacts of a commonly used active learning teaching strategy, group work, on student performance, using meta-analysis. We collected data from 34 articles and calculated estimates for each experiment within the article. This resulted in 66 estimates that represent 18,494 students. Our overall estimate indicates the implementation of group work in biology classrooms increased student performance by 0.72 standard deviations. When applied to a grade scale based on real-world data, the magnitude of this change is roughly one letter grade. Moderator analyses revealed the increase in performance holds across all class sizes, class levels, and group sizes up to five students. These results demonstrate group work leads to impressive boosts in student performance and reveal the value of studying specific active learning strategies.

In Chapter 3, I investigated student perceptions of and their preparation habits for online open-note exams in an undergraduate biology class, as compared to their previous experiences with closed-note exams in other classes. Specifically, I explored the following research questions: (1a) How do students perceive open-note exams impact their exam scores, their anxiety, the amount they studied, and the amount their peers studied? (1b) How do these perceptions impact performance outcomes? (2a) How do students prepare for open-note exams? (2b) How do these preparation methods impact performance outcomes? Results demonstrate students perceived increased exam scores, decreased exam-anxiety, decreased study time spent personally, and decreased study time spent by their peers for open-note exams, as compared to past experiences with closed-note exams. Open-ended survey responses analyzed through first- and second-cycle analyses showed students adapted their study habits by focusing on note preparation and broad conceptual understanding rather than rote memorization. Using linear mixed effects models to assess student performance, we found students who focused on understanding, note preparation and using external resources outperformed students who did not report those study habits. As institutions shift towards flexible and scalable assessments that can be used in face-to-face or online environments, the use of open-note exams can promote effective study habits and reward higher-order thinking with intentional guidance from the instructor.

In Chapter 4, I documented how the emergency transition to remote-online instruction, in response to the COVID-19 pandemic, impacted undergraduate students studying the principles of evolution in an introductory organismal biology class over time; specifically, how their study habits for exams differed (a) one week and (b) one month after a university's decision to transition to emergency remote instruction. I asked students about the extent to which COVID-19 impacted their study habits, and we categorized students’ responses using open coding. I identified several consistent similarities—as well as dramatic differences—in their responses as the time away from campus increased, documenting barriers and recommendations based on literature concerning crises and equitable practices.

Taken together, my work demonstrates both course elements and course structure have real impacts on student affect and performance. This dissertation defines active learning for the field of biology education research, evaluates the effect of the most used active learning strategy in biology education (i.e., group work) on student performance in undergraduate biology courses from multiple research articles written by multiple institutions, demonstrates study habits for open-note exams impact student performance on those exams, and documents challenges students faced during the pandemic. Further, it adds additional nuance to the discussion of active learning, student-centered strategies, evidence-based practices, and course elements.

REFERENCES

Adams, P.E., Driessen, E.P., Enya Granados, Isabell Wagner, Penny Ragland, Abby E. Beatty & Cissy Ballen. (In review). Embracing the inclusion of societal concepts in science leads to better understandings of their impacts. *Frontiers in Education*.

Beatty, A. E., Ballen, C. J., Driessen, E. P., Schwartz, T. S., & Graze, R. M. (2021b). Addressing the unique qualities of upper-level biology course-based undergraduate research experiences through the integration of skill-building. *Integrative and comparative biology*, *61*(3), 981-991.

Beatty, A. E., Driessen, E. P., Gusler, T., Ewell, S., Grilliot, A., & Ballen, C. J. (2021a). Teaching the tough topics: fostering ideological awareness through the inclusion of societally impactful topics in introductory biology. *CBE—Life Sciences Education*, *20*(4), ar67.

Beatty A.E., Driessen, E.P., Clark, A.D., Costello, R.A., Ewell, S., Fagbodun, S., Klabacka, R.L., Lamb, T., Mulligan, K., Henning, J.A., & Ballen, C.J. (In revision). Biology instructors see value in discussing controversial topics but fear personal and professional consequences. *CBE-LSE.*

Tracy, C. B., Driessen, E. P., Beatty, A. E., Lamb, T., Pruett, J. E., Botello, J. D., ... & Ballen, C. J. (2022). Why students struggle in undergraduate biology: Sources and solutions. *CBE—Life Sciences Education*, *21*(3), ar48.

Appendix A

Categories Defined

Category	Definition
Assessment (A)	This bin collects any strategies that helps the teacher evaluate students' knowledge/achievement in the class via feedback.
Conceptual Class Design (C)	This bin collects any strategies that concern how a teacher designed their class and values they hope to project. This bin focuses primarily on long term methods of affecting the classroom.
Discussion (D)	This bin collects any strategies that include people talking to one another
Games (G)	This bin collects any strategies that are "a form of play or sport, especially a competitive one played according to rules and decided by skill or luck" (Merriam Webster, 2019).
Group Work (W)	This bin collects any strategies that involve 2-6 students working together (Miller & Tanner, 2015).
Live-Action Visuals (V)	This bin collects any strategies used in class for student observation.
Meta-Cognition (M)	According to Flavel ("Metacognition and cognitive monitoring" 1979), the mental process of acquiring knowledge and understanding of cognitive phenomena through thought, experience, and the senses. i.e. oral communication of info, oral persuasion, oral comprehension, reading comprehension, writing, language acquisition, attention, memory, problem solving, social cognition, various types of self-control and self- instruction
Paperwork (K)	This bin collects any strategies that involve distributed handouts to be filled out by the student(s) that are not assumed to be graded.
Practicing Core Competencies (P)	This bin collects any strategies that fall under one or more of the six core competencies put forward by Vision and Change: (1) Apply the Process of Science (PPS); (2) Use quantitative reasoning (PQR); (3) Use Modeling and Simulation (PMS); (4) Interdisciplinary Nature

	(PIN); (5) Communicate and Collaborate (PCC); and/or (6) Relate Science and Society (PSS; AAAS, 2010).

Appendix B

Active Learning Strategy Guide

2 minute essays | M | A brief active-learning strategy that provides a mechanism for students to stop, think, and write during or at the end of a class period. The goal is to provide a momentary break during which students can capture their thoughts or questions (Miller & Tanner, 2015).

Synonyms: Minute papers, Notecards, & One-minute papers

5-E learning cycle | C | "This module provides a built-in structure for creating a constructivist classroom: the 5E Instructional Model. The 5E model sequences learning experiences so that students have the opportunity to construct their understanding of a concept over time. The model leads students through five phases of learning that are easily described using words that begin with the letter E: Engage, Explore, Explain, Elaborate, and Evaluate"(Inquiry, the Learning Cycle, & the 5E Instructional Model, NIH).

ABCs of Learning | C | A book presenting a new science of learning so that educators can creatively translate the science into exceptional practice. With everyday language, engaging examples, a sense of humor and solid evidence, it describes 26 unique ways that students learn (Schwartz, Tsang, & Blair, 2016).

Act out chromosome movements during meiosis | V, A | Pretending to be a chromosome during the process of meiosis to explain meiotic processes. Found in Newman and Wright (2017).

Active listening | M | Active listening means fully concentrating on what is being said rather than just passively ‘hearing’ the message of the speaker.

Active reading | M | Active reading is a process or technique of actively engaging with the text being read.

Analyses/presentation of literature | PCC | Students performing detailed examination/presentation of the elements or structure of primary literature. Found in Hoskins et al (2011).

Analyze data | PQR | Detailed examination/presentation of the elements or structure of data. Found in Escobar et al. (2019); Schultheis & Kjelvik (2015).

Annotating figures | M, PCC | Asking students to explain or comment on a figure. Found in Round & Campbell (2013).

Application of science and reasoning skills | PPS | An activity that focuses on having students apply scientific concepts and reasoning skills.

Applying knowledge of other subjects | PIN | An activity that focuses on having students apply their knowledge in other disciplines to a scientific problem. Found in Kohn et al. (2018).

Assessments | A | Assessments are any activity that gives feedback to the teacher about the student(s) understanding and mastery of material. Depending on the instructional goals, classroom assessment techniques (CATs) can employ closed-ended assessment questions, open-ended assessment questions, and questions at a variety of Bloom's taxonomy levels (Miller & Tanner, 2015).

Synonyms: CAT Scans

Audience response questions | A | A classroom response system is a set of hardware and software that facilitates teaching activities by collecting student answers and giving real time feedback to the teacher. Found in Caldwell (2007); Knight et al. (2013); Smith et al. (2009); Smith et al. (2011).

Synonyms: Student Response Systems, Clicker Questions, Personal Response System, iClicker, Classpulse, Classroom Polling, Electronic Polling, & Live Polling

Bingo | G | Review game in which students try to answer questions on a game board similar to non-academic versions of the game. Found in Tinsley (2019).

Brainstorming | M | Idea generation focused on a topic or question.

Break down scientific arguments | PPS, M | Analyzing and understanding a scientific argument.

Building models | PMS | Conceptually devise a representation of a system, object, or phenomenon. Not assumed to be a physical task. Found in Wilson et al. (2020).

Canvas collaboration | W, D | Use of the online classroom system Canvas for collaborative work.

Card sorts | A, D, M | This strategy gives students the opportunity to work with vocabulary, terms, and concepts. Students sort cards with the terms and concepts into categories based on meaning. Sorting the cards gives students a structure to talk meaningfully with one another about content and helps teachers check for understanding (Smith et al., 2013).

Carousel graffiti | D, M | A Carousel Activity is a communicative and interactive opportunity for participants/students to get up and move around a room in a circular fashion (much like a real carousel might do), stopping intermittingly to comment, discuss, or respond (verbally or in writing) to probing headings/ questions/topics/themes posted by a facilitator/teacher that is related to a given topic/theme.

Case studies | PSS | Case-based learning begins with a situation or scenario that poses one or more issues the students need to address. Importantly, both case-based and problem-based learning usually involve students working in structured groups in which they collaboratively identify questions and confusions and seek out additional information to expand their understanding of concepts related to the problem or case (Miller & Tanner, 2015). Found in Allen & Tanner (2005); Emtage et al. (2016); Herreid (1994); Stowell & Martin (2016).

Circulate to check for understanding | A, C | Activities pursued by teachers to keep track of student learning for purposes of making instructional decisions and providing feedback to students on their progress.

Class activities | O | Classroom activities are those done by student inside the class as part of applying or doing the practical part of the lesson after listening to the theoretical part which is presented by an instructor.

Class discussion | D | A classroom discussion is a sustained exchange between and among teachers and their students with the purpose of developing students' capabilities or skills and/or expanding students' understanding—both shared and individual—of a specific concept or instructional goal.

Cold call | A | Calling on a student whose hand is not raised. Freeman. also Knight et al. 2013.

Collaboration | PCC | Science is, by nature, a collaborative endeavor, and all scientific careers to which undergraduate students aspire will require extensive skills in working collaboratively. Group work, also referred to as cooperative learning, is a term that refers to activities that require students to engage in active learning with others, during which they work together toward a common outcome and practice improving their collaborative skills (Johnson et al., 1998). Group work and cooperative learning is often assumed to include more than two students, but usually no more than six (Miller & Tanner, 2015).

Collecting data | PPS | The collection of data through experimentation and/or observation. Found in Parrotta et al. (2019).

Compare & contrast | D, M | Teacher and student(s) comparing and contrasting concepts or scenarios to fuel discussion.

Competitive quizzes | A, G | A game-like function to increase student engagement while reviewing.

Complete exercises | O | Completion of exercises.

Completing charts | PQR | Teacher and students completing charts.

Completing supporting activities when assessments reveal a problem area | C | Completing supporting activities when assessments reveal a problem area.

Computer-based summative measure of student performance for each objective | A | Computer based summative measure of student performance for each objective.

Concept maps | M, V | Drawings or diagrams showing the mental connections that students make between a major concept the instructor focuses on and other concepts they have learned (Angelo & Cross, 2012).

Construct phylogenetic trees | M | Students constructing phylogenetic trees. Used in Ballen & Greene (2017); Eddy et al. (2013); Karimi et al. (2017).

Counter stereotypical scientist activity | M | Metacognitive assignments to help traditionally underserved students persist and succeed in science. Found in Schultheis & Kjelvik (2015); Schinske et al. (2016); Project Biodiversify: www.projectbiodiversify.org.

Creating skits | V | Skits to help understand and demonstrate content.

Creative activities | M | Activities to help understand content and generate higher level thinking.

Critical thinking exercises | M | Exercises that involve critical thinking about course material. Found in Styers et al. (2018).

Crossword puzzles | G | Games to help remember content and build verbiage familiarity.

Dancing | V | Dancing for learning purposes.

Data activities | PQR | Using data from experimentation/observation in an activity. Found in Schultheis & Kjelvik (2015).

Data collection | PPS | Collecting data from an experiment/observation. Found in Hanauer et al. (2017).

Debates | D | A regulated discussion of a proposition between two matched sides. Found in Vandegrift & Dawson (2016).

Debrief | M | To carefully review upon completion for the purpose of building greater understanding

Demonstrations | V | An act, process, or means of demonstrating knowledge; an outward expression or display.

Design experiments | PPS | Designing an experiment to test a hypothesis.

Designing vignettes | V | Creating a picture or sketch.

Development of physical props | V | Developing props to better explain content.

Diagramming | M, V | Creation of a diagram.

Discuss literature | D, PCC | Discussion of primary literature.

Discussion questions | D | Questions to spark discussion.

Discussions | D | Consideration of a question in open and usually informal debate.

Dissections | PPS, V | Dissecting an animal in a lab setting.

Drawing | V | Illustrating concepts visually in an informal, non-schematic representation.

Essay outlines | M | Creating the basic outline of an essay, helping to organize ideas.

Ethical stakeholder activities | PSS | Activities demonstrating how your decisions and desires as a stakeholder have implications. Found in Larson & Wong (2019).

Evaluate data | PPS | An activity in which students are asked to draw conclusions from a data set.

Exam redux | M | Review of exam material after the exam has been taken.

Exam Wrappers | M | Exam wrappers are short activities that direct students to review their performance (and the instructor\'s feedback) on an exam with an eye toward adapting their future learning. Exam wrappers ask students three kinds of questions: How did they prepare for the exam? What kind of errors did they make on the exam? What could they do differently next time? Each of the question types is discussed next (Carnegie Mellon University, 2020). Found in Smith et al. (2019).

Exercises that lead students to draw their own conclusions | M | Exercises that lead students to draw their own conclusions.

Exit tickets | A | Before students leave, they hand the teacher a ticket filled out with an answer to a question, a solution to a problem, or a response to what they've learned. (Lansing Community College Web Management, 2008)

Experimental design | PPS | Taking the time and effort to organize the experiment properly to ensure that the right type of data, and enough of it, is available to answer the questions of interest as clearly and efficiently as possible

Experiments | PPS | A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact.

Explanations | O | Making known or understandable.

Family feud | G | Review game loosely following the same structure as the television show of the same name.

Feedback | A | The transmission of evaluative or corrective information about an action, event, or process to the original or controlling source.

Field trips | V | Often significantly enhance the content of a course by providing information that is difficult to convey in a laboratory or classroom setting.

Fix the scenario | M | Student(s) work to identify problems and propose solutions to a scenario given by the teacher.

Flipped classroom | C | Students gain first-exposure learning prior to class and focus on the processing part of learning (synthesizing, analyzing, problem-solving, etc.) in class (Tucker, 2012). Found in Jensen et al. (2015); van Vliet et al. (2015).

Flow charts | M, V | A diagram that represents a process or chain of thought.

Follow instructions | C | Following instructions.

Formative assessment | A | This diagnostic use of assessment to provide feedback to teachers and students over the course of instruction is called formative assessment (Boston, 2002).

Gallery walk | D, W | This discussion technique allows students to be actively engages as they walk in team throughout the classroom. They work together in small groups to share ideas and respond to meaningful questions, documents, images, problem-solving situations or texts (Francek, 2006).

Games | G | Review activities designed to help students learn or retain material in an engaging and fun way.

Garage demos | V | Based on experience of students in introductory physics, where large-scale in-class demonstrations had been very effective in solidifying understanding and ability to recall basic principles, teachers use common items to create models and demonstrations to engage students.

Grant proposals | PPS, PSS, PCC | A formal proposal submitted to a government or civilian entity that outlines a proposed project and shows budgetary requirements and requests monetary assistance in the form of a grant

Graphic organizers/worksheets | M, V | A visual representation of knowledge, a way of structuring information, and of arranging essential aspects of an idea or a topic into a pattern using labels.

Graphing | PQR | Creating graphs based on data.

Group activities | W, D | Science is, by nature, a collaborative endeavor, and all scientific careers to which undergraduate students aspire will require extensive skills in working collaboratively. Group work, also referred to as cooperative learning, is a term that refers to activities that require students to engage in active learning with others, during which they work together toward a common outcome and practice improving their collaborative skills (Johnson et al., 1998). Group work and cooperative learning is often assumed to include more than two students, but usually no more than six (Miller & Tanner, 2015).

Synonyms: Cooperative Learning Activities

Group assignments | GW, D | Group work, also referred to as cooperative learning, is a term that refers to activities that require students to engage in active learning with others, during which they work together toward a common outcome and practice improving their collaborative skills (Johnson et al., 1998). Group work and cooperative learning is often assumed to include more than two students, but usually no more than six (Miller & Tanner, 2015).

Group brainstorming | W, M | Working together in a group of 2-6 students to propose solutions to a problem or to answer a question

Group discussion | GW, D | Discussion in a group of 2-6 students (Miller & Tanner, 2015).

Group discussion questions | W, D, A | Discussion in a group of 2-6 students aimed at providing a solution to a question (Miller & Tanner, 2015).

Group exams | W, D, A | Collaboration on an exam in a group of 2-6 students (Miller & Tanner, 2015). Found in Cortright et al. (2003).

Group learning | W, D | Science is, by nature, a collaborative endeavor, and all scientific careers to which undergraduate students aspire will require extensive skills in working collaboratively. Group work, also referred to as cooperative learning, is a term that refers to activities that require students to engage in active learning with others, during which they work together toward a common outcome and practice improving their collaborative skills (Johnson et al., 1998). Group work and cooperative learning is often assumed to include more than two students, but usually no more than six (Miller & Tanner, 2015).

Synonyms: Cooperative Learning

Group problems | W, D | Working on a problem as a group of 2-6 students.

Group projects | W, D, A, M | Science is, by nature, a collaborative endeavor, and all scientific careers to which undergraduate students aspire will require extensive skills in working collaboratively. Group work, also referred to as cooperative learning, is a term that refers to activities that require students to engage in active learning with others, during which they work together toward a common outcome and practice improving their collaborative skills (Johnson et al., 1998). Group work and cooperative learning is often assumed to include more than two students, but usually no more than six (Miller & Tanner, 2015).

Group random call | W, D, A | A group of 2-6 students are randomly selected to answer a question/problem posed by the teacher. Found in Knight et al. (2013).

Group work | W, D | Science is, by nature, a collaborative endeavor, and all scientific careers to which undergraduate students aspire will require extensive skills in working collaboratively. Group work, also referred to as cooperative learning, is a term that refers to activities that require students to engage in active learning with others, during which they work together toward a common outcome and practice improving their collaborative skills (Johnson et al., 1998). Group work and cooperative learning is often assumed to include more than two students, but usually no more than six (Miller & Tanner, 2015).

Group work with worksheets | W, D, K | Work given to a group of 2-6 students that includes or centers around worksheets (Miller & Tanner, 2015).

Groups Develop Hypothesis | W, D, PPS | To work out the possibilities of a tentative assumption made in order to draw out and test its logical or empirical consequences as a group.

Guest Lectures | C | A discourse given before an audience or class especially for instruction presented by someone who is not the assigned lecturer.

Guided Self-Directed Learning | C | Directed for or by oneself knowledge acquisition that is exhibited and explained to highlight points of interest.

Hands-on Activities | M | Characterized by active personal involvement an educational procedure designed to stimulate learning by firsthand experience

Hands-on Computer work | M, PMS | To perform or carry through a task requiring sustained effort or continuous repeated operations on a computer characterized by active personal involvement.

Hands-on laboratory experiments | M, PMS | An operation or procedure carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law characterized by active personal involvement.

Homework Assignments | A | To appoint as a task preparatory reading or research.

Hybrid CURE | C, PPS | Course-based undergraduate research experiences that involve whole classes of students in addressing a research question or problem that is of interest to the scientific community blended with more traditional lecture design.

Hypothesis Generation | PPS | To bring into existence a tentative assumption made in order to draw out and test its logical or empirical consequences.

Identifying Students' Misconceptions | M | To establish the distinguishing character of a wrong or inaccurate idea held by students.

iFAT | A | Immediate Feedback Assessment Technique: Students take a multiple-choice test in which they can immediately see if the answer they gave was right or wrong (Smith, J. A. 2013). Found in Cotner et al. (2008).

Synonyms: iRAT

In-class work | M | To perform or carry through a task requiring sustained effort or continuous repeated operations in the period during which such a body meets.

In-class writing assignments | M | Assignments given in class that serve to have students think and write about some aspect of the course.

Inclusive teaching | C | To conduct instruction regularly, allowing and accommodating for people who have historically been excluded (as because of their race, gender, sexuality, or ability). Found in Dewsbury & Brame (2019).

Index card activity | A, D, M | An educational procedure designed to stimulate learning by having students write questions they have about the course material on index cards, some of which will be discussed at the end of class.

Individual brainstorming | M | The mulling over of ideas by one person in an attempt to devise or find a solution to a problem.

Individual problem sets | M, A | Intended for one person, a group of a specific number of repetitions of a particular exercise, questions raised for inquiry, consideration, or solution.

Individual projects | A, M | Intended for one person, a task or problem engaged in usually by a group of students to supplement and apply classroom studies.

Individual reflections | M | Intended for one person, a period of consideration of some subject matter, idea, or purpose.

Individual work | M | To perform or carry through a task requiring sustained effort or continuous repeated operations intended for one person.

Individual writing | M | A written composition intended for one person to complete.

Inquiry | M | A systematic investigation and an examination into facts or principles. Found in Allen & Tanner (2005); Frisch et al. (2018); National Research Council (1996); Uno (1990).

Interactive and constructive activities | C | A mutually or reciprocally active educational procedure designed to stimulate learning by firsthand experience that promotes improvement or development.

Interactive games | G | Any mutually or reciprocally active activity undertaken or regarded as a contest involving rivalry, strategy, or struggle.

Interactive lectures | D | A mutually or reciprocally active discourse given before an audience or class especially for instruction.

Interactive tutorials | V | A mutually or reciprocally active paper, book, film, or computer program that provides practical information about a specific subject.

Interactive videos | V, D | A mutually or reciprocally active recording of a motion picture.

Interpret data | PPS | To explain or tell the meaning of or to present in understandable terms the factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation.

Interpret figures/graphs | PPS | To explain or tell the meaning of or to present in understandable terms a diagram or pictorial illustration of textual matter or a diagram that represents the variation of a variable in comparison with that of one or more other variables.

Jeopardy! | G, A | A quiz competition in which contestants are presented with clues in the form of answers and must phrase their responses in the form of questions (Benek-Rivera & Mathews, 2004).

Jigsaw | W, D, M | Divide students into small groups, giving each group a problem. Ask students to work to formulate effective ways of teaching problem-solving strategies for that kind of problem to another group of students. After students have had a chance to master their problem, bring the groups back together so that each group has an expert in each problem type (Collaborative Learning Techniques). Found in Theobald et al. (2017); Wiggins et al. (2017).

Journal club | W, D, M | An association of persons usually meeting periodically sharing records of experiences, ideas, or reflections.

Just-in-time telling | C | Also known as Just-in-time teaching, is a teaching and learning strategy comprised of both classroom activities that promote active learning and web-based resources that are used to enhance the classroom activities (Novak et al., 1999). Found in Marrs & Novak (2004).

Kahoot | G, A | Kahoot! is a game-based learning platform, used as educational technology in schools and other educational institutions. Its learning games, "Kahoots", are multiple-choice quizzes that allow user generation and can be accessed via a web browser or the Kahoot app (Kahoot!, 2013)

Knowledge checks | A, M | To compare with a source, original, or authority a student's acquaintance with or understanding of a science, art, or technique.

Lab | PPS | An academic period set aside for laboratory work.

Lab notebook | M | A book for notes or memoranda detailing experiences and knowledge gained in a place equipped for experimental study in a science or for testing and analysis.

Learning Catalytics | W, D, A, M | Engage your students in team-based and group-learning activities with this interactive student response tool (Pearson, 2011).

Learning Goals | A, M | Setting goals for students to work toward that will give them a chance to think on their understanding of the material and show their progress. Found in Allen & Tanner (2006).

Literature review | M | An activity that uses critical evaluation of the body of writings on a particular subject.

Low risk pre lecture quizzes | A | A short oral or written test not likely to result in failure, harm, or injury administered before the period of instruction. Found in Ballen et al. (2017); Casper et al. (2019); Moravec et al. (2010).

Make graphs | PPS | To put together from components a diagram (such as a series of one or more points, lines, line segments, curves, or areas) that represents the variation of a variable in comparison with that of one or more other variables.

Make predictions | M | To lay out and construct a foretelling on the basis of observation, experience, or scientific reason.

Making models | PPS | To lay out and construct a system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs.

Making quiz questions | M | An activity where students create short oral or written test questions.

Manipulables | V | Items that a student can move, arrange, operate, or control with their hands or another body part or by mechanical means, especially in a skillful manner (Krech, 2000).

Many hands, many voices | A | After asking a question, ask for multiple hands and multiple voices to respond to any question posed during class time to broaden participation and increase the breadth of ideas flowing from students to instructors (Tanner, 2013).

Synonyms: "Multiple hands, multiple voices"

Metacognition activities | M | An educational procedure designed to stimulate awareness or analysis of one's own learning or thinking processes by firsthand experience. Found in Dye & Stanton (2017); Flavell (1979); Sabel et al. (2017); Tanner (2012).

Mini lectures | C | Lectures that last 15 min or less to clarify concepts and highlight terminology from the textbook Found in Walker et al. (2008).

Model building | PPS | The formation and subsequent development of mental models by a learner (Buckley, 2012).

Model-based learning | M | The formation and subsequent development of mental models by a learner (Buckley, 2012).

Synonyms: Models

Module | C | An educational unit which covers a single subject or topic

Muddiest point | M | The technique consists of asking students to jot down a quick response to one question: “What was the muddiest point in [the lecture, discussion, homework assignment, film, etc.]?” The term “muddiest” means “most unclear” or “most confusing” (Mcdaniel, 2018).

Multiple choice conceptual questions | A | Examination into an abstract or generic idea generalized from instances with more than one pre-written answer to select from.

Normalize error | C | Make it clear to students that incorrect answers are just as fundamental to education as correct answers and are completely normal (Puryear, 2014).

Not Lecture | C | Not a discourse given before an audience or class especially for instruction

Note-checking | M, W, D | Students to compare their notes with a partner. This exercise allows students to see how other students take notes, and it also gives students the quick opportunity to reconsider what was important in the material.

Observing Cellular Mechanisms | PMS | To come to realize or know especially through consideration of noted facts.

Online learning modules | C | An educational unit which covers a single subject or topic available on the internet. Found in Goff et al. (2017).

Online quizzes | A | A short oral or written test administered online. Found in Brown et al. (2015); Pan et al. (2019).

Online textbook | C | A textbook available on the internet.

Open-ended questions | D, A, M | Permit spontaneous and unguided responses.

Synonyms: Oral questions; Open-ended questions about potential complications related to research, Short answer discussion questions

Pair discussions | W, D | Consideration of a question in open and usually informal debate by two or more students.

Paper discussions | D | Consideration of a question in open and usually informal debate on published work.

Participatory projects | W, D | A task or problem engaged in usually by a group of students to supplement and apply classroom studies.

Pause and recall | D, W, M | Short pauses built into lectures in which students are given a chance to recall what they had learned so far in lecture and to review it with their peers (Harrington, 2014).

Peer discussion | W, D | Consideration of a question in open and usually informal debate between students.

Peer evaluation | A | Determination of the value, nature, character, or quality of one's work by another student. Found in Smith et al. (2009).

Peer instruction | C, W, D | Students teach each other (Rubin & Hebert, 1998). Found in Snyder et al. (2016).

Synonyms: Peer teaching

Peer review | A | The active learning strategy of peer review was used to enhance student understanding and engagement in the critique process. This active learning strategy involved small groups of students who worked together as a team to evaluate the work of other student groups using a critique-rubric (Odom, Glenn, Sanner, & Cannella, 2009).

Perform analyses | PQR | A detailed examination of anything complex in order to understand its nature or to determine its essential features: a thorough study.

Plays and videos | V | (1) the stage representation of an action or story; (2) being, relating to, or involving images on a television screen or computer display.

POGIL | W, D, N, C | POGIL is an acronym for Process Oriented Guided Inquiry Learning. Because POGIL is a student-centered instructional approach, in a typical POGIL classroom or laboratory, students work in small teams with the instructor acting as a facilitator. The student teams use specially designed activities that generally follow a learning cycle paradigm. These activities are designed to have three characteristics (POGIL, 1994). Found in Murray (2014); Roller & Zori (2017).

Poll everywhere | A | A website where an instructor can pose a question and students respond simultaneously from the privacy of their tablets or other devices. Since feedback is anonymous, everyone can focus on what is said rather than who said it. All students have an equal opportunity to be heard (Poll Everywhere, 2007).

Poster presentation | D, A, PCC | A poster presentation is the presentation of research information in the form of a paper poster that conference participants may view (Purrington, 2010).

Practice exams | M | To perform an exercise designed to examine progress or test qualification or knowledge repeatedly so as to become proficient.

Practice questions | M | An interrogative expression often used to test knowledge to perform at repeatedly so as to become proficient.

Practicing skills | PPS | To perform the ability to use one's knowledge effectively and readily in execution or performance repeatedly so as to become proficient.

Praise effort | C | Verbally validate student effort (Tennessee Behavior Supports Project at Vanderbilt, 2016).

Praise improvement | C | Explicitly recognizing and praising student growth (Becker et al., 2017).

Pre-class reading and writing | C | Reading and writing assignments to be done by students prior to the material being covered in class. Found in

Pre-class tasks | C | A piece of work to be finished before the next class period.

Synonyms: Student preparation before class

Preceptors | C | Tutors.

Synonyms: Tutors

Predict-explain-observe | PPS | POE is a strategy often used in science. It works best with demonstrations that allow immediate observations and suits Physical and Material World contexts. A similar strategy also works well in mathematics, particularly in statistics. It can be used for: finding out students' initial ideas, providing teachers with information about students' thinking, generating discussion, motivating students to want to explore the concept, generating investigations (White & Gunstone, 1992).

Predicting cellular mechanisms | PPS | To declare or indicate in advance; hypothesize how they work.

Presentation to Class | A, M | Asking students to present material to their peers (with their peers involved in the assessment) would alert all the students to: ‘what makes a good presentation’; increase their subject-specific knowledge; and reinforce their immediate (and the wider) group’s identity (Dart, 2006). Found in Allen & Tanner (2005); Eisen (1998).

Problem-based learning | C, W, M | Student learning is promoted by the introduction of complex, real-world problems (Woods, 1985). Found in Allen & Tanner (2003); Tatner & Tierney (2016).

Problem sequences | M | A continuous or connected series of questions raised for inquiry, consideration, or solution.

Synonyms: problem sets

Problem solving | M | To find a solution, explanation, or answer for questions raised for inquiry, consideration, or solution.

Synonyms: Solving problems

Project-based learning | C, M | A model that organizes learning around projects. Projects are complex tasks, based on challenging questions or problems, that involve students in design, problem-solving, decision making, or investigative activities; give students the opportunity to work relatively autonomously over extended periods of time; and culminate in realistic products or presentations (Thomas, Mergendoller, & Michaelson, 1999). Found in Wright & Boggs (2002); Zwick (2018).

Projects | A, M | A planned piece of work that has a specific purpose (such as to find information or to make something new) and that usually requires a lot of time. : a task or problem in school that requires careful work over a long period of time.

Propose further research | PPS | To set forth studious inquiry or examination for acceptance or rejection.

Puzzles | G | A question, problem, or contrivance designed for testing ingenuity.

Quantitative problems | PPS | As for a quantitative problem, it requires the student to manipulate a formula or work through an algorithm to find a numerical solution to the problem

(Nakhleh, 1993). Found in Aikens & Dolan (2014); Andrews et al. (2017); Bray et al. (2016); Goldstein & Flynn (2011).

Question & answer | D, A | a period of time when people can have their questions answered. Synonyms: Students answering questions

Questioning prior knowledge | M | To subject the range of one's information or understanding to analysis.

Quiz | A | A short oral or written test.

Quiz bowl | A, G | Quiz bowl is a game in which two teams compete head-to-head to answer questions from all areas of knowledge (National Academic Quiz Tournaments, LLC, 2019).

Random call | A | The instructor may select students from the class to participate in specific activities. For smaller-sized courses, students may be selected from a cup holding popsicle sticks labeled with each student's name. For larger course class sizes, instructors can choose index cards, labeled with each student's name, out of a stack student (Tanner, 2012). Found in Ballen et al. (2019); Eddy et al. (2014). Knight et al 2013.

Reading | M | To learn from what one has seen or found in writing or printing.

Reading and evaluating scientific literature | M | To learn from what one has seen or found in writing or printing and to determine or fix the value of it. Found in Colosi & Zales (1998); Murray (2014).

Reading assignments | M, PPS | The reading of a passage assigned by the teacher.

Reading graphs | PPS, M | The interpretation of a visual representation of data/results (Tairab & Khalaf Al-Naqbi, 2004).

Redistribution of point allocation | C | The instructor reworks their grading system to reward group work and ongoing preparation rather than exam performance exclusively (Ballen, Wieman, Salehi, Searle, & Zamudio, 2017).

Reflect on experience | M | Consideration of some subject matter, idea, or purpose.

Synonyms: reflections, reflecting on the effectiveness of your study habit, reflecting on cellular mechanisms, reflective pauses, reflective writing

Reflective index cards | M | Students write out one to three ideas on an index card that capture their initial thoughts on how to answer a question posed by the instructor. The act of writing itself may lead students to discover points of confusion or key insights. If collected, this writing can hold students accountable in thinking and recording their ideas (Tanner, 2013).

Reflective journals | M | Prior to class, students complete assigned readings and then write about what they learned (e.g. key points, summary, understanding, discovery, and points of confusion). The students then participate in class activities that cover the topics from the readings. After class, the students revise the pre-class reflective journal in another color font and submit it for credit as the post-class reflective journal (Long, Su, & Waugh, 2013).

Relating scientific concepts to everyday phenomena or human experiences | PSS | The relation of the biological concepts learned in class to the real world (AAAS, 2010).

Synonyms: Think of science within the context of society

Reporting cellular mechanisms | PCC |

Research projects | PPS | A scientific investigation, usually using scientific methods, to achieve defined objectives. Found in Bakshi et al. (2019); Cotner & Hebert 2016.

Research skills | PPS | Developed aptitude or ability for diligent searching.

Researching information | PPS | Careful or diligent search.

Respond to instructor questions | A | To say something in return.

Results generation | PPS | Something obtained by calculation or investigation.

Retaking exams | A | A chance for students to take an assessment again.

Reviewing exam questions | M | The instructor explains assessment questions to the class. By recognizing their errors, each student can gain a better understanding (Hassan, 2011).

Role-playing | V | To represent in action.

Running simulations | PMS | The imitative representation of the functioning of one system or process by means of the functioning of another. Found in Bergan-Roller et al. (2017).

Synonyms: Simulations

Science argumentation | PCC | Students form persuasive exchanges to construct knowledge about the natural world and science (Munford, 2002). Found in Lacum et al. (2014).

Science skills | PCC ALL | A variety of developed attributes that aid in the processes or understanding of science. According to AAAS (2010), there are six core competencies every undergraduate biology student should practice and hone.

Self-directed learning | M | The overt management of the external learning environment (Pilling-Cormick & Garrison, 2007).

Self-reflections | M | An examination of one's own thoughts and feelings.

Sentence sorts | W, M | Below are three different ways to use sentence strips in BrainPOP ELL lessons. They all involve cognitive skills – putting items on a continuum, sequencing events from the movie, and matching sentence halves. Using manipulative is always good because students are actively engaged in the activity. Do them as partner activities, so students are using the target vocabulary and concepts as they share ideas and communicate with each other.

Sequence reconstructions | M | “Analyze and depict graphically a series of events, actions, roles or decisions. Useful for understanding processes, cause and effect, and chronological series, and organizing information in an orderly, coherent progression” (Barkley, Cross, & Major, 2005 p. 206).

Set up an experiment | PPS |

Synonyms: Set up and maintain their own aquaria

Share prior knowledge | D, PCC | Students reveal their current understanding of a particular topic (Meyrick, 2011).

Synonyms: Sharing info

Skits | V | Either (1) a satirical or humorous story or sketch or (2) a short serious dramatic piece especially one done by amateurs.

Snowball responses | M, D, W | Learning cooperative with snowball throwing method to train students to be more responsive independently and enable more intensive interaction occur both in asking questions and opinions with other students in one group in the form of a snowball made of paper (Marlena, 2016).

Socratic discussion | M, D | The act of teaching did not consist in transmitting information from "teacher" to "student," but was an exercise in helping students to cease their reliance on perceptual knowledge of the imperfect material world, and stimulating them to introspectively discover true knowledge through logic and reasoning. prompting students, through cross-examination, into acknowledging their own fallacies and then asking them provocative questions to steer them towards realizing true knowledge via introspection (Stoddard & O’Dell, 2016).

Synonyms: Socratic method, Socratic questioning, Thought questions

Sorting steps | M, W | Below are three different ways to use sentence strips in BrainPOP ELL lessons. They all involve cognitive skills – putting items on a continuum, sequencing events from the movie, and matching sentence halves. Using manipulative is always good because students are actively engaged in the activity. Do them as partner activities, so students are using the target vocabulary and concepts as they share ideas and communicate with each other.

Stand and shout outs | A | The instructor asks a question to which each individual student will respond. Each response is usually less than 30 seconds in length (Tanner, 2013).

Synonyms: Whip around

Statement corrections | K | The instructor provides statements, readings, proofs, or other material that contains errors. The students are charged with finding and correcting the errors. Concepts that students commonly misunderstand are well suited for this activity (The Derek Bok Center for Teaching and Learning at Harvard University, 2016)

Stretch-it questions | M | A technique in which the instructor asks additional questions, after asking an initial question, to extend the response and the learning even further. Often, questions such as, “How? and “Why?” are repeatedly asked (Lemov, 2010).

Strip sentences | M | The goal of this activity is for students to order a set of items, such as steps in a biological process or a series of historical events. As one strategy, the instructor provides students with a list of items written on strips of paper for the students to sort (Bhatt, 2007).

Synonyms: Strip sequences

Structured problems | M | Follow a structured format to solve problems. It is useful for "dividing problem-solving processes into manageable steps so that students don't feel overwhelmed and so that they learn to identify, analyze, and solve problems in an organized manner" (Barkley, Cross, & Major, 2005, p.171)

Student ownership | C | Students are charged with leading their learning journey, and the teacher becomes a guide to aid them on this journey (Weimer, 2002).

Synonyms: Student-centered; Supporting students working individually

Student-led discussion | D, M | Students enrolled in the class each take a turn in facilitating a discussion with a group of five to seven peers on an assigned article. The facilitator's responsibility is to review the article, prepare questions that he/she believes would spark discussion, and facilitate the group discussion. The role of the group members who are not facilitators that day is to read the article and come prepared with a typed sheet of three points that were of interest to them, and about which they are prepared to speak (McGlynn-Stewart, 2015).

Summarizing | M | Covering the main points succinctly.

Synthesis | M | The combining of often diverse concepts into one coherent concept.

Teachers answering questions | C | Teachers answering questions posed by the students

Team-based learning quizzes | W, A | The completion of a short oral or written test by a group of students.

Synonyms: Team quizzes, Team-based learning

Team challenges | W, M | The completion of a stimulating task or problem by a group of students.

Synonyms: Team activities, Small-group active-engagement (GAE) exercises

Team-based learning | C | Students use class concepts to solve problems within their groups. The students must be held accountable for their individual and group work and provided with frequent feedback. The assigned groups need to promote learning and team interdependence (Michaelsen & Sweet, 2008). Found in Haidet et al. (2014); Jeno et al. (2017).

Test feedback with discussion | D, M | Getting tests back, looking over answers, discussing them and what you got wrong and why.

Think out loud | M, W | Learners have opportunities to talk with their peers about their ideas, observations, and understanding. These opportunities are helpful because in expressing their thinking out loud, learners: (1) find out what they understand and don't understand; (2) make connections between new and existing knowledge; and (3) reflect on their own thinking.

Think-pair-share | W, D, M | Involves giving all students a minute or so to think (or write) about their ideas concerning a biological question. Students are then tasked with turning and talking with a neighboring student to compare ideas and identify points of agreement and misalignment (Tanner, 2013).

Synonyms: Turn and talk

Thinking time | M | An allotted period to use one's mind to produce thoughts.

Top hat | C | Top Hat's active learning technology helps professors engage students before, during, and after class. This engagement includes access to electronic textbooks, assessments, assignments, and in class feedback (Top Hat, 2009).

Troubleshoot | M | Troubleshooting exercises have been commonly included in the problem sections of recently published textbooks on electronics and circuit analysis. Such problems demand a minimum level of knowledge and comprehension that has to be applied before any conclusion(s) may be analyzed then synthesized in order to facilitate the evaluation of the resulting outcome(s) by the student (Banky & Wong, 2007).

Two-stage exams | W, A, M | Students take the same exam repeatedly. They first take it individually, then take it again working in pairs, and then finally take it for the last time while working in larger groups (Yuretich, Khan, Leckie, & Clement, 2001). This approach uses the exam itself not only for evaluation, but also as a learning tool (Zipp, 2007). Found in Wieman et al. (2014).

Using models | PMS | To utilize a representation of something.

Venn diagrams | M, V | A graph that employs circles to represent logical relations between the terms of propositions by the inclusion, exclusion, or intersection of the curves.

Verbal questions | A | A spoken interrogative expression often used to test knowledge.

Videos | V | A digital recording of an image or set of images (e.g. movie or animation).

Weekly review | C | A study of material previously studied, conducted once every seven-day period.

Working problems | M | The manner of functioning or operating a question raised for inquiry, consideration, or solution.

Working with models | PMS | The manner of functioning or operating a miniature representation of something.

Worksheets | K | A sheet of paper on which are printed exercises and problems to be solved by a student.

Writing | M | A letter, note, or notice used to communicate or record.

Synonyms: Writing exercises, writing prompts, writing time

Writing assignments | M, A | A specified task or amount of work assigned or undertaken as if assigned by authority.

The following strategies were not defined or categorized because we either (1) did not know what they meant or (2) they were vague: ASK, CREATE, Real science, Study learning, Talk to text, Throat vote, and White board activities.

‘

Appendix C

References Used to Define Active Learning Strategies

AAAS (2010). Vision and Change: A Call to Action accessed 11 March 2010 Washington, DC AAAS www.visionandchange.org/VC_report.pdf. Google Scholar

Aikens, M. L., & Dolan, E. L. (2014). Teaching quantitative biology: Goals, assessments, and resources. Molecular Biology of the Cell, 25(22), 3478-3481.

Allen, D., & Tanner, K. (2003). Approaches to cell biology teaching: learning content in context—problem-based learning. Cell Biology Education, 2(2), 73-81.

Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex. Cell biology education, 4(4), 262-268.

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. CBE—Life Sciences Education, 5(3), 197-203.

Angelo, T. A., & Cross, K. P. (2012). Classroom assessment techniques. Jossey Bass Wiley.

Andrews, S. E., Runyon, C., & Aikens, M. L. (2017). The math–biology values instrument: Development of a tool to measure life science majors' task values of using math in the context of biology. CBE—Life Sciences Education, 16(3), ar45.

Bakshi, A., Webber, A. T., Patrick, L. E., Wischusen, W., & Thrash, C. (2019). The CURE for Cultivating Fastidious Microbes. Journal of microbiology & biology education, 20(1).

Ballen, C. J., Aguillon, S. M., Awwad, A., Bjune, A. E., Challou, D., Drake, A. G., ... & Harcombe, W. (2019). Smaller Classes Promote Equitable Student Participation in STEM. BioScience, 69(8), 669-680.

Ballen, C. J., & Greene, H. W. (2017). Walking and talking the tree of life: Why and how to teach about biodiversity. PLoS biology, 15(3).

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. CBE—Life Sciences Education, 16(4), ar56.

Banky, George P., and Kwong K. Wong. (2007). "Troubleshooting exercises using circuit simulator software: support for deep learning in the study of electronic circuits." International Conference on Engineering Education, Coimbra, Portugal.

Barkley, E. F., Cross, K. P., & Major, C. H. (2005). Collaborative learning techniques, San Francisco: JosseyBass.

Becker, E. A., Easlon, E. J., Potter, S. C., Guzman-Alvarez, A., Spear, J. M., Facciotti, M. T., ... & Pagliarulo, C. (2017). The effects of practice-based training on graduate teaching assistants' classroom practices. CBE—Life Sciences Education, 16(4), ar58.

Benek-Rivera, J., & Mathews, V. E. (2004). Active Learning with Jeopardy: Students Ask the Questions———. Journal of Management Education, 28(1), 104-118.

Bergan-Roller, H.E., Galt, N.J., Dauer, J.T., and Helikar, T. 2017. Discovering Cellular Respiration with Computational Modeling and Simulations. CourseSource. https://doi.org/10.24918/cs.2017.10

Bhatt, R. (2007). Modern Methods of Teaching Biology. APH Publishing.

Boston, C. (2002). The concept of formative assessment. Practical Assessment, Research, and Evaluation, 8(1), 9.

Bray, S.R., Duffin, P.M., and Wagner, J.D. (2016). Thinking deeply about quantitative analysis: Building a Biologist's Toolkit. CourseSource. https://doi.org/10.24918/cs.2016.4

Brown, G. A., Bice, M. R., Shaw, B. S., & Shaw, I. (2015). Online quizzes promote inconsistent improvements on in-class test performance in introductory anatomy and physiology. Advances in physiology education, 39(2), 63-66.

Buckley, B. C. (2012). Model-based learning. Encyclopedia of the sciences of learning, 5, 2300-2303.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. CBE—Life Sciences Education, 6(1), 9-20.

Carnegie Mellon University. (2020). Exam Wrappers. Retrieved from https://www.cmu.edu/teaching/designteach/teach/examwrappers/

Casper, A. M., Eddy, S. L., & Freeman, S. (2019). True Grit: Passion and persistence make an innovative course design work. PLoS biology, 17(7), e3000359.

Colosi, J. C., & Zales, C. R. (1998). Jigsaw cooperative learning improves biology lab courses. Bioscience, 48(2), 118-124.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W. & DiCarlo, S. E. Student retention of course content is improved by collaborative-group testing. Adv. Physiol. Educ.27, 102–108 (2003).

Cotner, S. H., Fall, B. A., Wick, S. M., Walker, J. D., & Baepler, P. M. (2008). Rapid feedback assessment methods: Can we improve engagement and preparation for exams in large-enrollment courses?. Journal of Science Education and Technology, 17(5), 437-443.

Cotner, S., & Hebert, S. (2016). Bean beetles make biology research sexy. The American Biology Teacher, 78(3), 233-240.

Dart, J. (2006). Developing a learning environment conducive to active learning and participation: Group presentations and formative assessment at level one. Journal of Hospitality, Leisure, Sport and Tourism Education, 5(1), 58-65.)

Derek Bok Center for Teaching and Learning at Harvard University, 2016). Active Learning. Retrieved February 2020, from https://bokcenter.harvard.edu/active-learning

Dewsbury, B., & Brame, C. J. (2019). Inclusive teaching. CBE—Life Sciences Education, 18(2), fe2.

Dye, K. M., & Stanton, J. D. (2017). Metacognition in upper-division biology students: Awareness does not always lead to control. CBE—Life Sciences Education, 16(2), ar31.

Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. CBE—Life Sciences Education, 13(3), 478-492.

Eddy, S. L., Crowe, A. J., Wenderoth, M. P., & Freeman, S. (2013). How should we teach tree-thinking? An experimental test of two hypotheses. Evolution: education and outreach, 6(1), 13.

Eisen, A. (1998). Small group presentations—teaching "science thinking" and context in a large biology class. BioScience 48,53 -58.

Emtage, L., Bradbury, L., Coleman, N., Devenport, D., Nietel, A., and Grew, J. (2016). Cell Signaling Pathways: A case study approach. CourseSource. https://doi.org/10.24918/cs.2016.9

Escobar, M.A., Morgan, W., Makarevitch, I., and Robertson, S.D. (2019). Tackling "Big Data" with Biology Undergrads: A Simple RNA-seq Data Analysis Tutorial Using Galaxy.CourseSource. https://doi.org/10.24918/cs.2019.13

Flavell JH (1979). Metacognition and cognitive monitoring: a new area of psychological inquiry. Am Psychol 34, 906–911.

Francek, M. (2006). Promoting discussion in the science classroom using gallery walks. Journal of College Science.

Frisch, J. K., Jackson, P. C., & Murray, M. C. (2018). Transforming undergraduate biology learning with inquiry-based instruction. Journal of Computing in Higher Education, 30(2), 211-236.

Goff, E. E., Reindl, K. M., Johnson, C., McClean, P., Offerdahl, E. G., Schroeder, N. L., & White, A. R. (2017). Efficacy of a meiosis learning module developed for the virtual cell animation collection. CBE—Life Sciences Education, 16(1), ar9.

Goldstein, J., & Flynn, D. F. (2011). Integrating active learning & quantitative skills into undergraduate introductory biology curricula. The American Biology Teacher, 73(8), 454-461.

Haidet, P., Kubitz, K., & McCormack, W. T. (2014). Analysis of the team-based learning literature: TBL comes of age. Journal on excellence in college teaching, 25(3-4), 303.

Hanauer, D. I., Graham, M. J., Betancur, L., Bobrownicki, A., Cresawn, S. G., Garlena, R. A., ... & Jacobs, W. R. (2017). An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. Proceedings of the National Academy of Sciences, 114(51), 13531-13536.

Harrington, C. (2014). The Power of Pausing. Retrieved February 13, 2020, from https://blog.cengage.com/power-pausing/

Hassan, O. A. (2011). Learning theories and assessment methodologies–an engineering educational perspective. European Journal of Engineering Education, 36(4), 327-339.

Herreid, C.F. (1994). Case studies in science: a novel method of science education. J. College Sci. Teaching 23(4),221 -229. Google Scholar

Hoskins, S. G., Lopatto, D., & Stevens, L. M. (2011). The C.R.E.A.T.E. Approach to Primary Literature Shifts Undergraduates' Self-Assessed Ability to Read and Analyze Journal Articles, Attitudes about Science, and Epistemological Beliefs. CBE—Life Sciences Education. 10(401).

Jeno, L. M., Raaheim, A., Kristensen, S. M., Kristensen, K. D., Hole, T. N., Haugland, M. J., & Mæland, S. (2017). The relative effect of team-based learning on motivation and learning: a self-determination theory perspective. CBE—Life Sciences Education, 16(4), ar59.

Jensen JL, Kummer TA, Godoy PDdM (2015). Improvements from a flipped classroom may simply be the fruits of active learning. CBE Life Sci Educ 14, ar5. Link, Google Scholar

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works?. Change: the magazine of higher learning, 30(4), 26-35.

Kahoot! (2013, March). Learning Games: Make Learning Awesome! Retrieved February 21, 2020, from https://kahoot.com/

Karimi, N., Parks, B.M., Rouse, D., Martin, K., Dong, X., Rajangam, P.C., Baum, D.A., and Heitz, J.G. (2017). Building Trees: Introducing evolutionary concepts by exploring Crassulaceae phylogeny and biogeography. CourseSource.https://doi.org/10.24918/cs.2017.16

Krech, B. (2000). Math: Model with Manipulatives. Instructor, 109(7), 6.

Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: student reasoning and the impact of instructional cues. CBE—Life Sciences Education, 12(4), 645-654.

Kohn, K. P., Underwood, S. M., & Cooper, M. M. (2018). Connecting Structure–Property and Structure–Function Relationships across the Disciplines of Chemistry and Biology: Exploring Student Perceptions. CBE—Life Sciences Education, 17(2), ar33.

Lacum, E. B. V., Ossevoort, M. A., & Goedhart, M. J. (2014). A teaching strategy with a focus on argumentation to improve undergraduate students' ability to read research articles. CBE—Life Sciences Education, 13(2), 253-264.

Lansing Community College Web Management. (n.d.). Center for Teaching Excellence. Retrieved from https://www.lcc.edu/cte/resources/teachingtips/tip33.aspx

Larson, E.I. and Wong, M.Y. (2019). Using structured decision making to explore complex environmental issues. CourseSource. https://doi.org/10.24918/cs.2019.18

Lemov, D. (2010). Teach like a champion: 49 techniques that put students on the path to college (K-12). John Wiley & Sons.

Long, T., Su, C., & Waugh, M. (2013). Using A Flipped-Classroom Instructional Model in A Large-Enrollment Undergraduate Genetics Class: An Action Research Study. In Knoxville. The University of Tennessee-Knoxville. The Annual Convention of the Association for Educational Communications and Technology (Vol. 1, No. 1, pp. 109-116).

Marlena, N. (2016). Implementation of learning snowball throwing method to improve student learning outcomes of operation management subject. In Proceeding The 1st International Conference on Economics Education & Entrepreneurship (pp. 238-242).

Marrs, K. A., & Novak, G. (2004). Just-in-time teaching in biology: Creating an active learner classroom using the internet. Cell Biology Education, 3(1), 49-61.

Mcdaniel, R. (2018). Classroom Assessment Techniques (CATs). Retrieved February 21, 2020, from https://cft.vanderbilt.edu/guides-sub-pages/cats/

McGlynn-Stewart, M. E. (2015). Undergraduate Students' Perspectives on the Value of Peer-Led Discussions. Canadian Journal for the Scholarship of Teaching and Learning, 6(3), 9.

Meyrick, K. M. (2011). How STEM education improves student learning. Meridian K-12 School Computer Technologies Journal, 14(1), 1-6.

Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. New directions for teaching and learning, 2008(116), 7-27.

Miller, S., & Tanner, K. D. (2015). A portal into biology education: an annotated list of commonly encountered terms. CBE—Life Sciences Education, 14(2), fe2.

Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before lecture: a strategy that improves learning outcomes in a large introductory biology class. CBE—Life Sciences Education, 9(4), 473-481.

Munford, D. (2002). Situated argumentation, learning and science education: A case study of prospective teachers’ experiences in an innovative science course.

Murray, T. A. (2014). Teaching students to read the primary literature using POGIL activities. Biochemistry and Molecular Biology Education, 42(2), 165-173.

Newman, D.L., and Wright, L.K. (2017). Meiosis: A Play in Three Acts, Starring DNA Sequence. CourseSource. https://doi.org/10.24918/cs.2017.9

Nakhleh, M. B. (1993). Are our students conceptual thinkers or algorithmic problem solvers? Identifying conceptual students in general chemistry. Journal of Chemical Education, 70(1), 52.

National Academic Quiz Tournaments, LLC. (2019). What is quiz bowl? Retrieved February 13, 2020, from https://www.naqt.com/

National Research Council (Ed.). (1996). National science education standards. Washington: National Academy Press.

Novak, G., Patterson, E., Gavrin, A., & Christian, W. (1999). Just-in-time teaching: Blending active learning with web technology. Upper Saddle River, NJ: Prentice Hall

Odom, S., Glenn, B., Sanner, S., & Cannella, K. A. (2009). Group Peer Review as an Active Learning Strategy in a Research Course. International Journal of Teaching and Learning in Higher Education, 21(1), 108-117.

Pan, S. C., Cooke, J., Little, J. L., McDaniel, M. A., Foster, E. R., Connor, L. T., & Rickard, T. C. (2019). Online and Clicker Quizzing on Jargon Terms Enhances Definition-Focused but Not Conceptually Focused Biology Exam Performance. CBE—Life Sciences Education, 18(4), ar54.

Parrotta MD, Dickinson J, Merritt TJS. (2019). Gotcha! Which fly trap is the best? An introduction to experimental data collection and analysis. CourseSource. https://doi.org/10.24918/cs.2019.35.

Pearson. (2011, July). Learning Catalytics. Retrieved February 21, 2020, from https://www.pearson.com/us/higher-education/products-services-teaching/learning-engagement-tools/learning-catalytics.html

Pilling-Cormick, J., & Garrison, D. R. (2007). Self-directed and self-regulated learning: Conceptual links. Canadian Journal of University Continuing Education, 33(2).

POGIL. (1994). What is POGIL? Retrieved February 13, 2020, from https://pogil.org/about-pogil/what-is-pogil.

Poll Everywhere. (2007). Classroom discussions without anxiety. Retrieved February 2020, from https://www.polleverywhere.com/k12-student-response-system

Purrington, Colin (2010). Designing Conference Posters. Retrieved February 13, 2020, from https://colinpurrington.com/tips/poster-design/

Puryear , J. (2014, May). What Do Great Teachers Do? The Techniques that Separate Great Teachers from Average Teachers . Retrieved February 21, 2020, from https://www.thedialogue.org/wp-content/uploads/2015/04/PolicyBrief-GreatTeachers-FINAL.pdf

Roller, M. C., & Zori, S. (2017). The impact of instituting process-oriented guided-inquiry learning (POGIL) in a fundamental nursing course. Nurse education today, 50, 72-76.

Round, J. E. & and A. Malcolm Campbell, A. M. (2013). Figure Facts: Encouraging Undergraduates to Take a Data-Centered Approach to Reading Primary Literature. CBE—Life Sciences Education. 101(12).

Rubin, L., & Hebert, C. (1998). Model for active learning: Collaborative peer teaching. College Teaching, 46(1), 26-30.

Sabel, J. L., Dauer, J. T., & Forbes, C. T. (2017). Introductory biology students' use of enhanced answer keys and reflection questions to engage in metacognition and enhance understanding. CBE—Life Sciences Education, 16(3), ar40.

Schinske, J. N., Perkins, H., Snyder, A., & Wyer, M. (2016). Scientist spotlight homework assignments shift students' stereotypes of scientists and enhance science identity in a diverse introductory science class. CBE—Life Sciences Education, 15(3), ar47.

Schultheis, E. H., & Kjelvik, M. K. (2015). Data nuggets: bringing real data into the classroom to unearth students' quantitative & inquiry skills. The American Biology Teacher, 77(1), 19-29.

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them. WW Norton & Company.

Smith, B. A., Metzger, K., & Soneral, P. (2019). Investigating Introductory Nonmajor Biology Students' Self-Regulated Learning Strategies Through the Implementation of a Reflective-Routine. Journal of College Science Teaching, 48(6), 66-76.

Smith, J. A. (2013). Immediate Feedback and Assessment Technique (IF-AT) testing forms: An overview of the tool and uses. In Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL conference (Vol. 40).

Smith, M. K., Trujillo, C., & Su, T. T. (2011). The benefits of using clickers in small-enrollment seminar-style biology courses. CBE—Life Sciences Education, 10(1), 14-17.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. Science, 323(5910), 122-124.

Snyder, J. J., Sloane, J. D., Dunk, R. D., & Wiles, J. R. (2016). Peer-led team learning helps minority students succeed. PLoS biology, 14(3).

Stoddard, H. A., & O’Dell, D. V. (2016). Would Socrates have actually used the “Socratic Method” for clinical teaching?. Journal of general internal medicine, 31(9), 1092-1096.

Stowell, S.M., and Martin, A.P. (2016). Cutthroat trout in Colorado: A case study connecting evolution and conservation. CourseSource. https://doi.org/10.24918/cs.2016.20

Styers, M. L., Van Zandt, P. A., & Hayden, K. L. (2018). Active learning in flipped life science courses promotes development of critical thinking skills. CBE—Life Sciences Education, 17(3), ar39.

Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs?. Journal of Biological Education, 38(3), 127-132.

Tanner, K. D. (2012). Promoting student metacognition. CBE—Life Sciences Education, 11(2), 113-120.

Tanner, K. D. (2013). Structure matters: twenty-one teaching strategies to promote student engagement and cultivate classroom equity. CBE—Life Sciences Education, 12(3), 322-331.

Tatner, M., & Tierney, A. (2016). An extended, problem-based learning laboratory exercise on the diagnosis of infectious diseases suitable for large level 1 undergraduate biology classes. Journal of Biological Education, 50(1), 54-60.

Tennessee Behavior Supports Project at Vanderbilt. (2016). Behavior-Specific Praise in the Classroom. Retrieved February 13, 2020, from https://vkc.mc.vanderbilt.edu/assets/files/resources/psibehaviorspecpraise.pdf

Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, A. J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. PloS one, 12(7).

Thomas, J. W., Mergendoller, J. R., & Michaelson, A. (1999). Project based learning for middle school teachers. Middle School Journal, 36(2), 28-31.

Tinsley, H.N. (2019). Teaching cell structures through games. CourseSource. https://doi.org/10.24918/cs.2019.23

Top Hat. (2009). Top Hat, The Active Teaching Platform. Retrieved February 2020, from https://tophat.com/

Tucker, B. (2012). The flipped classroom. Education Next, 12(1), 82-83.

Uno, G.E. (1990). Inquiry in the classroom.BioScience40(11),841 -843.

Vandegrift, E.V.H., and Dawson, S.M. 2016. Sex and gender: What does it mean to be female or male? CourseSource. https://doi.org/10.24918/cs.2016.21

van Vliet EA, Winnips JC, Brouwer N (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. CBE Life Sci Educ 14, ar26. Link, Google Scholar

Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: integrating active learning into a large lecture course. CBE—Life Sciences Education, 7(4), 361-367.

Weimer, M. (2002). Learner-centered teaching: Five key changes to practice. John Wiley & Sons.

White, R. T., & Gunstone, R. F. (1992). Probing Understanding. Great Britain: Falmer Press.

Wieman, C. E., Rieger, G. W., & Heiner, C. E. (2014). Physics exams that promote collaborative learning. The Physics Teacher, 52(1), 51-53.

Wiggins, B. L., Eddy, S. L., Grunspan, D. Z., & Crowe, A. J. (2017). The ICAP active learning framework predicts the learning gains observed in intensely active classroom experiences. AERA Open, 3(2), 2332858417708567.

Wilson, K. J., Long, T. M., Momsen, J. L., & Speth, E. B. (2020). Modeling in the Classroom: Making Relationships and Systems Visible. CBE—Life Sciences Education 101 (19).

Woods, D. (1985). Problem-based learning and problem-solving. In: Problem-Based Learning for the Professions, ed. D. Boud. Sydney: HERDSA.

Wright, R., & Boggs, J. (2002). Learning cell biology as a team: a project-based approach to upper-division cell biology. Cell Biology Education, 1(4), 145-S27.

Yuretich, R. F., Khan, S. A., Leckie, R. M., & Clement, J. J. (2001). Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course. Journal of Geoscience Education, 49(2), 111-119.

Zipp, J. F. (2007). Learning by exams: The impact of two-stage cooperative tests. Teaching Sociology, 35(1), 62-76.

Zwick, M. (2018). The design, implementation, and assessment of an undergraduate neurobiology course using a project-based approach. Journal of Undergraduate Neuroscience Education, 16(2), A131.

Appendix D

Citations used by the articles that defined active learning

1. Smith M. K., Wood W. B., Adams W. K., Wieman C., Knight J. K., Guild N., Su T. T. (2009). Why peer discussion improves student performance on in-class concept questions. Science 323, 122-124.
2. Collins J. W., O’Brien N. P. (2003). The Greenwood dictionary of education, Eds. Westport, CT: Greenwood.
3. Haak D. C., HilleRisLambers J., Pitre E., Freeman S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. Science 332, (6034), 1213-1216 doi: 10.1126/science.1204820.
4. Freeman S., O’Connor E., Parks J. W., Cunningham M., Hurley D., Haak D., Wenderoth M. P. (2007). Prescribed active learning increases performance in introductory biology. CBE— Life Sciences Education 6, (2), 132-139 doi: 10.1187/cbe.06-09-0194.
5. Beichner R. J., Saul J. M., Abbott D. S., Morse J. J., Deardorff D., Allain R. J., Risely J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALEUP) project. Research-Based Reform of University Physics 1, (1), 2-39.
6. Lorenzo M., Crouch C. H., Mazur E. (2006). Reducing the gender gap in the physics classroom. American Journal of Physics 74, (2), 118-122.
7. Freeman S., Eddy S. L., McDonough M., Smith M. K., Okoroafor N., Jordt H., Wenderoth M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences of the USA 111, (23), 84108415 doi: 10.1073/pnas.1319030111.

8. Eddy S. L., Converse M., Wenderoth M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. CBE—Life Sciences Education 14, ar23 10.1187/cbe14- 06-0095.

9. Handelsman, J., Miller, S., & Pfund, C. (2007). Scientific teaching. New York: Macmillan.

10. Aragón, O. R., Dovidio, J. F., & Graham, M. J. (2016). Colorblind and multicultural ideologies are associated with faculty adoption of inclusive teaching practices. Journal of Diversity in Higher Education, 10(3), 10.1037/dhe0000026

11. Morse JM, Barrett M, Mayan M, Olson K, Spiers J (2002). Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods 1, 13-22.

12. Eddy SL, Brownell SE, Thummaphan P, Lan MC, Wenderoth MP (2015). Caution, student experience may vary: social identities impact a student's experience in peer discussions. CBE Life Sci Educ 14, ar45.

13. American Association for the Advancement of Science (AAAS) (2011). Vision and Change: A Call to Action, Washington, DC.

14. Brame, C. J. (2013). Flipping the classroom. Retrieved February 2, 2018, from https://cft.vanderbilt.edu/wp-content/uploads/sites/59/Flipping-the-classroom.pdf

15. Jensen, J. L., Kummer, T. A., & Godoy, P. D. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. CBE—Life Sciences Education, 14(1), ar5.

16. Cooper KM, Brownell SE (2016). Coming out in class: challenges and benefits of active learning in a biology classroom for LGBTQIA students. CBE Life Sci Educ 15, ar37.

17. Seidel SB, Reggi AL, Schinske JN, Burrus LW, Tanner KD (2015). Beyond the biology: a systematic investigation of noncontent instructor talk in an introductory biology course. CBE Life Sci Educ 14, ar43.

18. Eddy SL, Brownell SE, Wenderoth MP (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. CBE Life Sci Educ 13, 478-492.

19. Wiggins G., McTighe J. (2005). Understanding by design, Alexandria, VA: Association for Supervision and Curriculum Development.

20. Prince M. (2004). Does active learning work? A review of the research. Journal of Engineering Education 93, 223-231.

21. Frederick P. J. (1987). Student involvement: Active learning in large classes. New Directions for Teaching and Learning 1987, 45-56.

22. Armbruster P, Patel M, Johnson E, Weiss M (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. CBE Life Sci Educ 8, 203-213.

23. Freeman S, Haak D, Wenderoth MP (2011). Increased course structure improves performance in introductory biology. CBE Life Sci Educ 10, 175-186.

24. Singer S. R., Nielsen N. R., Schweingruber H. A. (2013). Biology education research: Lessons and future directions. CBE—Life Sciences Education 12, (2), 129-132.

25. Eberlein T, Kampmeier J, Minderhout V, Moog RS, Platt T, Varma-Nelson P, White HB (2008). Pedagogies of engagement in science: a comparison of PBL, POGIL, and PLTL. Biochem Mol Biol Educ 36, 262-273.

26. Sarquis JL, Dixon LJ, Gosser DK, Kampmeier JA, Roth V, Strozak VS, VarmaNelson P (2001, Ed. JE MillerJE GrocciaM Miller, The workshop project: peer-led team learning in chemistry In: Student-Assisted Teaching: A Guide to Faculty-Student Teamwork, Bolton MA: Anker, 150-155.

27. Lom, B. (2012). Classroom activities: simple strategies to incorporate studentcentered activities within undergraduate science lectures. J Undergrad Neurosci Educ. 11(1):A64–A71.

28. Chamany K, Allen D, Tanner K (2008). Making biology learning relevant to students: integrating people, history, and context into college biology teaching. CBE Life Sci Educ 7, 267278.

29. Greenwald AG, Banaji MR, Nosek BA (2015). Statistically small effects of the Implicit Association Test can have societally large effects. J Pers Soc Psychol 108, 553- 561.

30. Couch BA, Brown TL, Schelpat TJ, Graham MJ, Knight JK (2015). Scientific teaching: defining a taxonomy of observable practices. CBE Life Sci Educ 14, ar9.

31 Miller S, Pfund C, Pribbenow CM, Handelsman J (2008). Scientific teaching in practice. Science 322, 1329-1330. 31. Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM, et al. (2004). Scientific teaching. Science 304, 521- 522.

32. Wood WB, Handelsman J (2004). Meeting report: the 2004 National Academies summer institute on undergraduate education in biology. Cell Biol Educ 3, 215-217.

33. van Vliet EA, Winnips JC, Brouwer N (2015). Flipped-class pedagogy enhances student metacognition and collaborative-learning strategies in higher education but effect does not persist. CBE Life Sci Educ 14, ar26.

34. Stockwell BR, Stockwell MS, Cennamo M, Jiang E (2015). Blended learning improves science education. Cell 162, 933-936.

35. Strayer J (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. Learn Environ Res 15, 171-193.

36. Johnson DW, Johnson RT, Smith KA (1991). Cooperative Learning: Increasing College Faculty Instructional Productivity (ASHE-ERIC Higher Education Report No. 4), Washington, DC: George Washington University.

37. Crouch CH, Mazur E (2001). Peer instruction: ten years of experience and results. Am J Phys 69, 970-977.

38. Tanner KD (2009). Talking to learn: why biology students should be talking in classrooms and how to make it happen. CBE Life Sci Educ 8, 89-94.

39. Quillin K, Thomas S (2015). Drawing-to-learn: a framework for using drawings to promote model-based reasoning in biology. CBE Life Sci Educ 14, es2.

40. Reynolds JA, Thaiss C, Katkin W, Thompson RJ (2012). Writing-to-learn in undergraduate science education: a community-based, conceptually driven approach. CBE Life Sci Educ 11, 17-25.

41. Dori YJ, Belcher J (2005). How does technology-enabled active learning affect undergraduate students’ understanding of electromagnetism concepts. J Learn Sci 14, 243- 279.

42. Whiteside A, Brooks DC, Walker JD (2010). Making the case for space: three years of empirical research on learning environments. Educause www.educause.edu/ero/article/making-case-space-three-years-empirical-researchlearning-environments (accessed 22 August 2015).

43. Hibbard L, Sung S, Wells B (2016). Examining the effectiveness of a semi-selfpaced flipped learning format in a college general chemistry sequence. J Chem Educ 93, 24-30.

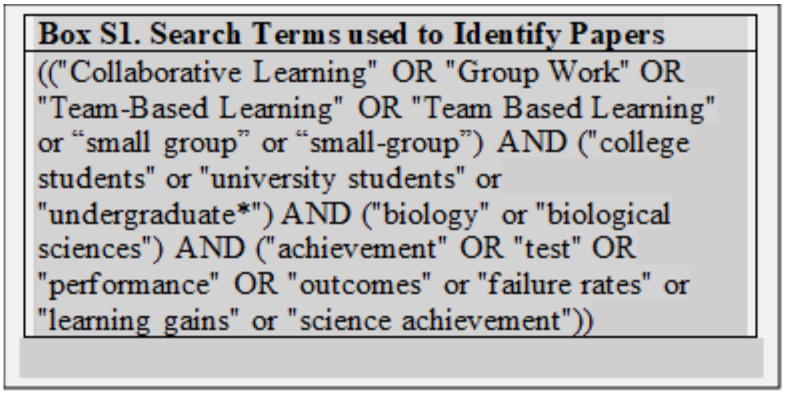
Appendix E

Meta-Analysis Extended Materials and Methods

Table S1. *Search specifics for each database*

Databases Searched	**Search Specifics**
Web of Science	Click on Advanced Search. Copy and paste search terms from Box 1 into the search box and then search for terms as topics (i.e., add “TS=” before the terms). Restrict language to English; Create a custom year range of 1924 to Nov, 16 2020.
PsycINFO	Click on Advanced Search. Copy and paste each set of search terms, separated by parentheses and an “AND” (e.g., “Collaborative Learning” OR “Group Work” OR “Team-Based Learning” OR “Team Based Learning” OR “small group” or “small-group”) into a separate search box. Choose “AND” in the dropdown boxes between search term lines. Select “TX All Text” for each field next to the search terms. Select “Boolean/Phrase for search modes. Check “Apply equivalent subjects.” Search from Jan 1924 to Nov 16, 2020. Check English.
ERIC (Ebsco)	Same as the instructions for PsycINFO.
PubMed	Click on Advanced Search. Copy and paste each set of search terms, separated by parentheses and an “AND” (e.g., “Collaborative Learning” OR “Group Work” OR “Team-Based Learning” OR “Team Based Learning” OR “small group” or “small-group”) into a separate search box. Choose “AND” in the dropdown boxes between search term lines. Select “Text Word” for each field next to the search terms. Press search. Then search from Jan 1924 to Nov 16, 2020.
Academic Search Premiere	Click on Advanced Search. Copy and paste each set of search terms, separated by parentheses and an “AND” (e.g., “Collaborative Learning” OR “Group Work” OR “Team-Based Learning” OR “Team Based Learning” OR “small

	group" or "small-group") into a separate search box. Choose "AND" in the dropdown boxes between search term lines. Select "AB Abstract or Author-Supplied Abstract" for each field next to the search terms. Select "Boolean/Phrase for search modes. Check "Apply equivalent subjects." Search from Jan 1924 to Nov 16, 2020. Select English. Check "Full text". Check "PDF Full Text"
Education Research Complete	Same as the instructions for Academic Search Premiere.



Box S1. Search Terms used to Identify Papers

(("Collaborative Learning" OR "Group Work" OR "Team-Based Learning" OR "Team Based Learning" or "small group" or "small-group") AND ("college students" or "university students" or "undergraduate*") AND ("biology" or "biological sciences") AND ("achievement" OR "test" OR "performance" OR "outcomes" or "failure rates" or "learning gains" or "science achievement"))

Data Collection: Choosing a Correlation Value

In the case of the paired design study, correlation values were not present in any of the studies, so we inputted a correlation value of 0.99 (i.e., the highest amount of correlation possible to enter into the CMA software) given that pre- and post-correlation is likely to be high given the same student population and performance measures were used before and after exposure to group work. In our dataset, the high correlation value resulted in nonpositive sampling variances (i.e., sampling variances of 0). This became an issue because the *metafor* package in R cannot compute heterogeneity values for the complete dataset if there are any nonpositive sampling variances. To correct this, we decreased the correlation value to 0.95, and we kept decreasing from that value by 0.05 until all our effect sizes had positive sampling variances. This brought us to the input correlation value of 0.90. This is ultimately the value we used to calculate our Hedges' *g* and sampling variance with for the paired study designs. However, we also conducted sensitivity analyses by changing our correlation values to 0.5 and 0.1, so the reader can see the effect of this on our dataset (Table S2F, S2G).

Data Collection: Missing values and Mean-Value imputation

When a study did not provide standard deviation/standard error values and/or the number of students in each group – the case for 8 of our 66 estimates (12.5%) – we estimated these values using the average standard deviation or average number of students from the other studies (Kambach et al., 2020). Since the studies we used fell into two different types of study design categories (i.e., independent or paired design) we used two different methods for calculating an average standard deviation:

To calculate an average standard deviation for studies with an independent design, we averaged standard deviations for the pre- and post-scores in the studies that fell into the independent design category. Specifically, two average standard deviation values were calculated: (1) average pre-score standard deviation; and (2) average post-score standard deviation. These calculations resulted in an average pre-score standard deviation value of 15.70 and an average post-score standard deviation value of 18.10. These values were used for one independent design study (Weasel & Finkel, 2016). Of note, we did not need to calculate an average student n for any of these independent design studies, however, we did calculate one for the paired design studies.

To calculate an average standard deviation and student n for paired design studies missing these values, we averaged standard deviations and student numbers for the paired design. Specifically, we calculated two average standard deviation values: (1) average treatment group standard deviation; and (2) average control group standard deviation. These calculations resulted in an average group standard deviation value of 14.68 and an average control standard deviation value of 16.26. These values were used for six different paired design studies representing seven estimates (Baines et al., 2004; Burrowes, 2003; Costa-Silva et al., 2018; Daniel, 2016; Huysken et al., 2019; Knight & Wood, 2005). Additionally, we calculated an average number of students for studies with a paired design. Specifically, we calculated two average student n values: (1) average number of treatment group students; and (2) average number of control group students. These calculations resulted in an average group student n value of 209 and an average control student n value of 144. These values were used for one paired design study representing one estimates (Baines et al., 2004).

Data Collection: The Case of Multiple Outcomes, Treatments, or Controls

We developed a protocol to handle cases where studies provided either multiple outcomes, multiple treatments, or multiple controls. When studies provided multiple outcomes from the same set of students, we chose the performance outcome with the largest number of student participants. If the reported outcomes all had the same number of student participants, then we constructed a hierarchy of selection: course grade>average exam score>final exam>midterm>concept inventory. Of note, course grade was only used in paired research designs if the grading schemes for both groups were the same. This hierarchy was used rather than averaging all the outcomes because in most cases, the data required to calculate effect size was not present or decipherable for all the outcome variables. Course grades were chosen over exams because course grades usually provided a more holistic picture than exams did, given there were often other course elements contributing to student grades, aside from just exams. Exams were chosen over concept inventories because students had a graded incentive to take them more seriously.

In some cases, studies provided data for multiple treatments or multiple controls. Since the studies fell into two different types of design categories (i.e., independent or paired design), we used two different methods to address these situations.

1. In the case of multiple treatments or multiple controls in an independent design, we utilized as much of the data provided as possible. For example, if there were three classes or courses with reported student performance outcomes after experiencing no group work (i.e., three control groups) and three classes or courses with reported student performance outcomes after experiencing group work (i.e., three treatment groups), then we pooled the mean, standard deviation and number of student participants using a pooling tool at the following website: https://home.ubalt.edu/ntsbarsh/business-stat/otherapplets/Pooled.htm. However, we only did this when the treatments were similar, and the controls were similar. If they were not similar, then we separated the different treatments with their similar controls and calculated multiple estimates rather than using one estimate.
2. In the case of a paired research design, there were no controls present, so only the case of multiple treatments sometimes applied. When this applied, we obtained as much data concerning each of the multiple treatments (i.e., mean student performance and standard deviation or standard error measures both before and after experiencing group work). In the case of missing data, we first wrote to the lead author, requesting the missing data. If we did not receive a response within 2 weeks, then, if possible, we pooled the treatment data together using a pooling tool at the following website:

https://home.ubalt.edu/ntsbarsh/business-stat/otherapplets/Pooled.htm. If pooling was not possible due to lack of data, then we dropped the specific treatment or paper from the study.

Coding Data: Categorizing class sizes.

We coded class size similarly to the procedure used in the Freeman et al. (2014) meta-analysis to keep our findings comparable to theirs (Freeman et al., 2014). Specifically, class size was treated as a continuous variable in all papers where it was mentioned. In cases where the authors reported a range of class sizes, we used the midpoint value in the range. Since Freeman et al. (2014) noted class sizes fell into three natural groupings (classes with 50 or fewer students, classes of more than 50 up to 110, and classes with more than 110), we used those three categories and designated them as small, medium, and large classes, respectively.

Coding Data: Categorizing group sizes.

We treated group size as a continuous variable in all papers where it was mentioned. In cases where the authors reported a range of group sizes, we used the midpoint value in the range. After collecting the data, we made a histogram of the group size distribution to create group size categories. To create categories with roughly equal distributions and frequencies, we decided on three natural groupings: groups with less than 4 students, groups with 4 and 5 students, and groups with more than 5 students. We designated these as small, medium, and large groups, respectively.

Appendix F

Meta-analysis Extended Results

Sensitivity analyses

We conducted several sensitivity analyses to determine the robustness of our results by considering the extent to which they are affected by changes in our models or assumptions. First, we removed the random effect of author to demonstrate inflation occurs without controlling for this variable (Table S2B). Then, because 26 of the 66 effect sizes in this meta-analysis came from one study (Weir et al., 2019), we analyzed the dataset without these effect sizes to test whether the effect of group work on student performance was not solely dependent on a single article (Table S2C). After the removal of these effect sizes, we found group work was still beneficial to student academic performance. We then conducted another analysis after the removal of an extreme outlier (i.e., ~25 standard errors larger than the mean; (Yapici, 2016)(Table S2D). To demonstrate that our decision to use the mean-value for standard deviation imputation is robust, we doubled the imputed standard deviation value (Table S2E). Of importance, none of these analyses significantly impacted our results or conclusions.

We also examined the data by the two types of research designs we considered in our meta-analysis: paired research designs (i.e., pre- versus post-group-work data collection) and independent research designs (i.e., post-treatment with group work versus post-control without group work). First, to test whether our dataset is robust to changes in the correlation value in paired research design comparisons, we changed the correlation value to 0.5 (Table S2F) and 0.1 (Table S2G). We then conducted a pairwise comparison of student performance as parsed out into paired and independent research designs. This comparison suggests the groups are statistically significantly different from each other (QM (df=1) = 4.53, P=0.03). More specifically, students in paired research designs performed 0.47 standard deviations (+/-0.43; +/- 95% C.I.) higher than the average student performance in independent research designs (p=0.03). To test whether studies with a paired design simply reflected performance gains over the duration of a course, we reanalyzed the dataset after excluding those data. We found the effect of group work was still positive and significant (Table S2H).

Together, these sensitivity analyses demonstrated the effect of group work on student performance is still statistically significant and robust to the removal of particularly large estimates, the removal of articles that contribute a large amount of estimates, the removal of estimates taken from paired research designs, inputted standard deviation changes, and correlation value changes.

Table S2: Sensitivity Analyses

Analysis	***n***	**Hedges' *g***	**s.e.**	**p-val**	**95% C.I.: lower limit**	**95% C.I.: upper limit**
All Estimates; author; 1 SD; cor 0.9	66	0.7162	0.1330	<0.0001	0.4555	0.9769
All Estimates; 1 SD; cor 0.9	66	0.9987	0.0062	<0.0001	0.9866	1.0108
Without Weir; author; 1 SD; cor 0.9	40	0.7071	0.1370	<0.0001	0.4385	0.9757
Without Yapici; author; 1 SD; cor 0.9	65	0.6367	0.1105	<0.0001	0.4201	0.8533
All Estimates; author; 2SD; cor 0.9	66	0.6447	0.1314	<0.0001	0.3872	0.9023
All Estimates; author; 1 SD; cor 0.5	66	0.7502	0.1418	<0.0001	0.4724	1.0281
All Estimates; author; 1 SD; cor 0.1	66	0.7523	0.1431	<0.0001	0.4718	1.0327
Independent Design Estimates, 1 SD	30	0.5902	0.1494	<0.0001	0.2973	0.8831

Assessing for Bias

To assess for bias, we conducted a visual analysis of the funnel plot including all 66 effect sizes. A funnel plot is a scatter plot of study effect sizes (x-axis) versus their estimated standard errors (y-axis); a smaller standard error equates to greater precision. A plot should be symmetrical, with increasing variability in effect sizes observed in the less precise studies towards the bottom of the plot, producing a funnel shape when no bias is present. Asymmetry in this plot may indicate that publication bias is present due to a lack of observed data points in a particular region of the plot (Sterne et al., 2002). Due to the perceived asymmetry of the funnel plot (Figure S1), we investigated publication bias further using Duval and Tweedie's trim and fill method (Duval & Tweedie, 2000b, 2000a). This method is used to estimate the number of studies missing from a meta-analysis due to the suppression of the most extreme results on one side of the funnel plot. The method then augments the observed data, yielding a more symmetric funnel plot. The trim and fill results estimated there were zero missing studies on the left side of the graph, making us less concerned about publication bias (Figure S2).

As a final check for publication bias, we calculated fail-safe N's using the Rosenthal method. This method calculates the number of studies with negative effect sizes that are necessary to change the significance level (p-value) to insignificant (≥ 0.05) (Rosenthal, 1979). The results demonstrated that changing the significance of our findings from $p<0.0001$ to $p=0.05$ would require 449,901 studies that yielded null results. Thus, these additional analyses add confidence that the meta-analysis is not significantly affected by publication bias.

Figure S1. Funnel plot of all 66 effect sizes before and after trim and fill. The trim and fill will add white dots where studies should be in the case of potential publication bias. However, this process estimated there were zero studies missing on the left side (SE =4.06).

Appendix G

References used in Appendix E & F

Baines, A. T., McVey, M., Rybarczyk, B., Thompson, J. T., & Wilkins, H. R. (2004). Mystery of the toxic flea dip: An interactive approach to teaching aerobic cellular respiration. *Cell Biology Education*, *3*(1), 062–068. https://doi.org/10.1187/cbe.03-06-0022

Burrowes, P. A. (2003). A student-centered approach to teaching general biology that really works: Lord's constructivist model put to a test. *The American Biology Teacher*, *65*(7), 491–502.

Costa-Silva, D., Côrtes, J., Bachinski, R., Spiegel, C., & Alves, G. (2018). Teaching Cell Biology to Dental Students with a Project-Based Learning Approach. *Journal of Dental Education*, *82*(03), 322–331. https://doi.org/10.21815/jde.018.032

Daniel, K. L. (2016). Impacts of Active Learning on Student Outcomes in Large-Lecture Biology Courses. *The American Biology Teacher*, *78*(8), 651–655. https://doi.org/10.1525/abt.2016.78.8.651

Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Duval, S., & Tweedie, R. (2000b). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

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

Huysken, K., Olivey, H., McElmurry, K., Gao, M., & Avis, P. (2019). Assessing Collaborative, Project-based Learning Models in Introductory Science Courses. *Journal of the Scholarship of Teaching and Learning*, *19*(1), 6–28. https://doi.org/10.14434/josotl.v19i1.26777

Kambach, S., Bruelheide, H., Gerstner, K., Gurevitch, J., Beckmann, M., & Seppelt, R. (2020). Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecology and Evolution*, *10*(20), 11699–11712.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(4), 298–310.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638.

Sterne, J. A. C., Jüni, P., Schulz, K. F., Altman, D. G., Bartlett, C., & Egger, M. (2002). Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological'research. *Statistics in Medicine*, *21*(11), 1513–1524.

Weasel, L. H., & Finkel, L. (2016). *Deliberative Pedagogy in a Nonmajors Biology Course: Active Learning That Promotes Student Engagement With Science Policy and Research*. http://www.nsta.org/college/

Weir, L. K., Barker, M. K., McDonnell, L. M., Schimpf, N. G., Rodela, T. M., & Schulte, P. M. (2019). Small changes, big gains: A curriculum-wide study of teaching practices and student learning in undergraduate biology. *PLoS ONE*, *14*(8), 1–16. https://doi.org/10.1371/journal.pone.0220900

Yapici, İ. Ü. (2016). Effectiveness of Blended Cooperative Learning Environment in Biology Teaching: Classroom Community Sense, Academic Achievement and Satisfaction. *Journal of Education and Training Studies*, *4*(4), 269–280. https://doi.org/10.11114/jets.v4i4.1372