Equity and inclusion underlie instructor assessment choices amid COVID-19 pandemic
by

Todd Lamb

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 7, 2022

Keywords: Emergency transition online, Diversity, Exams, Online Learning

Approved by

Cissy Ballen, Chair, Assistant Professor of Biological Sciences
Karen McNeal, Associate Professor of Geosciences
Paul Cobine, Professor and Chair of Biological Sciences

Abstract

The coronavirus (COVID-19) outbreak mandated a rapid transition to online classes with little warning. Previous literature studying the effect of this sudden shift demonstrated enormous impacts on instructors and students. However, the details concerning instructor assessment choices during this time are less clear. In chapter one of this work, we asked biology instructors to reflect on the changes they made to their assessments of student learning during the emergency transition to remote instruction in spring of 2020 and whether the potential changes were motivated by equity concerns. We also asked that instructors describe the assessment changes they intended to keep in future semesters. Through qualitative analyses, we found that instructors removed components of assessment more often than they added them, and the most common changes included how instructors administered exams and engaged students through participation. Instructors reported that equity concerns motivated their decision-making, particularly their concern over students' ability to access learning resources. Instructors indicated they would keep many of the changes they made in response to COVID-19. Our research shows the pandemic dramatically altered how instructors assessed students in biology, but equity-based decisions leading to lasting change may be one positive outcome for future students.

What is Discipline-Based Education Research?

Discipline-Based Education Research (DBER) is a collection of related research fields all in STEM. DBER researchers in physics, chemistry, engineering, biology, the geosciences, and astronomy study similar problems, use similar methods, and draw on similar theories and usually collaborate with each other to achieve goals successfully. The goals of DBER are to understand how people learn the concepts, practices, and ways of thinking of science and engineering;

understand the nature and development of expertise in a discipline, help identify and measure appropriate learning objectives and instructional approaches that advance students toward those objectives; contribute to the knowledge base in a way that can guide the translation of DBER findings to classroom practice; and identify approaches to make science and engineering education broad and inclusive (Singer et al., 2012).

In 2011, the American Association for the Advancement of Science with support from the National Science Foundation (NSF) met with faculty, administrators, students, and other stakeholders to discuss specifically, biology education research and how it could become more successful in future efforts. This discussion brought about the report "Vision and Change in Undergraduate Biology Education- A Call to Action" This report provided a set of principles to direct undergraduate biology education reform. This report also stipulated important guidance for best practices in pedagogy, the input of undergraduate students, and a lens for broadening participation and truly making biology inclusive of all students (American Association for the Advancement of Science, 2009).

Acknowledgments

First, I would like to thank Dr. Cissy Ballen for being such an amazing mentor. Her experience, expertise, and collaboration skills were key to my completion of this work and my degree as a whole. Dr. Ballen's constant support and drive continuously motivated me to stay on track and be as productive as possible while also enjoying this learning opportunity here at Auburn University. Her ability to connect me with others that work on this research was paramount in the completion of this project, as well as learning research techniques from many professionals nationally and internationally.

Additionally, I would like to thank the rest of the Ballen lab for their support and feedback throughout my time here as a graduate student. From research and classes, to professional development and job searching, the experience that other lab members were able to offer was essential to my present and future success. I want to give a special thanks to the undergraduate assistants who helped me process all of my data and work on numerous other side projects for this work that I could not have done alone.

I would also like to thank the collaborators of this project in the Equity and Diversity in Undergraduate STEM (EDU-STEM) research coordination network who helped with the revisions of this project at many different stages. Particularly Sehoya Cotner, Catherine Creech, Abby Grace Drake, Sheritta Fagbodun, Kristen S. Hobbs, A. Kelly Lane, Erin Larson, Sophie McCoy, and Seth Thompson. Graphs were created in R and many figures were formatted with Lucidapp.com. This work was supported by NSF DBI-1919462 awarded to Sehoya Cotner, Cissy Ballen, Sheritta Fagbodun, and the EDU-STEM research coordination network.

Finally, I would like to thank my family and friends for all of their support, specifically my fiancé who has worked with me through this time so that we could both be successful in our endeavors.

Table of Contents

List of Tables and Figures

List of Abbreviations

PEERS	Persons Excluded because of their Ethnicity or Race
STEM	Science, Technology, Engineering, and Math

CHAPTER ONE

Equity and inclusion underlie instructor assessment choices amid COVID-19 pandemic

Submitted for publication in *CBE-Life Sciences Education*

Todd Lamb[1], Emily P. Driessen[1], Abby E. Beatty[1], Rachel Youngblood[1], Abby Esco[1], Sehoya Cotner[2], Catherine Creech[3], Abby Grace Drake[4], Sheritta Fagbodun[5], Kristen S. Hobbs[6], A. Kelly Lane[7], Erin Larson[8], Sophie J. McCoy[9,10], Seth Thompson[7] & Cissy J. Ballen[1]

[1]Auburn University, Auburn, AL 36849

[2]University of Bergen, Bergen, Norway 5006

[3]Mt Hood Community College, Gresham, OR 97030

[4]Cornell University, Ithaca, NY 14850

[5]Tuskegee University, Tuskegee, AL 36088

[6]Kansas State University, Manhattan, KS 66506

[7]University of Minnesota Twin Cities, Minneapolis, MN 55455

[8]Alaska Pacific University, 4101 University Drive, Anchorage, AK 99508

[9]Florida State University, Tallahassee, FL 32306

[10]University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Introduction

Several teaching 'norms' have prevailed in undergraduate STEM classrooms for centuries, despite calls to replace outdated instructional strategies with evidence-based approaches. In a classic example, traditional lecture instruction is prominent throughout the undergraduate STEM curriculum (Patrick et al., 2016; Stains et al., 2018), despite overwhelming evidence that active learning leads to better and more equitable student outcomes (Ballen, Salehi & Cotner, 2017; Barral et al., 2018; Beichner et al., 2007; Freeman et al., 2007; Freeman et al., 2014; Haak et al., 2011; Lorenzo et al., 2006; Theobald et al., 2020; Wilton et al., 2019). Another persisting norm of STEM classrooms is the dominant or exclusive use of high-stakes, summative assessments (Goubeaud, 2010). However, previous research shows high-stakes testing is not an inclusive practice (Cotner and Ballen, 2017; Matz et al., 2017), nor is it a reliable reflection of student learning (Momsen et al., 2010).

Pedagogical change in higher education can be a slow-moving and complex process involving individuals, institutions, and policies (Reinholz et al., 2021). However, sometimes, an immediate need for change is triggered by external environmental factors that require instructors to quickly adapt their approaches. Such was the case when instructors were forced to transition their courses online after the start of the COVID-19 pandemic in the spring of 2020. This transition presented practical and complex obstacles for instructors (Gurung & Stone, 2020; Lashley et al., 2020), many of whom had little to no experience teaching online and were not offered support or training to aid their efforts (Lassoued et al., 2020). During this time, instructors changed multiple elements of their courses to adapt their face-to-face classes to the online remote environment. These changes affected all students, but particularly students from low socioeconomic backgrounds (Aucejo et al., 2020), students without access to adequate technology (Van Dijk, 2006), and students with computer anxiety (Brosnan, 1998). Instructor decisions about how to assess students in remote online environments had large impacts on student academic outcomes and their well-being (Nambiar, 2020).

Multiple factors may have influenced instructor choices about student assessments in spring 2020. For example, these decisions may have been driven by need-based factors, institutional requirements, concerns about equity, or other unforeseen factors. We reason that equity-driven decisions about assessments that instructors keep in subsequent semesters will represent one silver lining of the pandemic; but, if faculty made assessment changes because of need-based motivations, such as time-constraints, perceived limitations of online environments, or other 'necessary evils,' they may be less likely to keep those changes in the future. We define equity-driven decisions about assessments as those that focus on how different factors positively relate to advancements, gains, and excellence of all students (Pearson et al., 2022). While previous work demonstrates that professional development centered on equity issues help teachers create equitable spaces for students (Chittooran, 2020; Riordan et al., 2019), this is the first study, to our knowledge, that has focused on how these issues shape instructor decisions because of a single pervasive event. We conducted a systematic investigation into what specific changes instructors made to their assessments and what motivated those changes. Specifically, our research questions included: **(1)** What assessment changes, if any, did instructors implement during the emergency transition to online learning due to COVID-19? **(2)** Were any of these changes motivated by equity concerns? **(3)** How did potential changes in assessment due to COVID-19 impact instructor choices in subsequent semesters?

Methods

Instructor selection

In light of previous work that found biology education research was conducted disproportionately at large research universities (Thompson et al., 2020), we strove to diversify the institutions from which we drew a sample of participating instructors. We based our search for instructors across three institution types: (1) primarily white research-based institutions (doctoral-granting institutions and master's-granting institutions without a minority-serving designation; Carnegie Classification of Institutions of Higher Education, n.d.), (2) associate degree-granting colleges (institutions that offer training and classes that are affordable and relevant to the local community; Schinske et al., 2017; Carnegie Classification of Institutions of Higher Education, n.d.), and (3) minority serving institutions (a federally recognized category of establishment based on minority student enrollment criteria). We acknowledge that these categories can overlap (i.e., research intensive universities that are minority serving). However, we explicitly delineate between primarily white research-based institutions and the broader set of minority serving institutions because of the unique cultures associated with these different institutions, particularly with respect to institutional approaches to equity.

We randomly selected three institutions from each of these three categories per state in the United States (N = 9 institutions from each state). We selected these institutions using a random number generator. Then, we selected two undergraduate biology instructors at random from each institution, identified through publicly available information and university websites, and sent these individuals a link to our survey. The participant recruitment process is depicted in Figure 1. Some states did not have one or more of these institution types, and some institutions did not have an updated list of faculty with electronic contact information, so these were therefore excluded from the study population.

Through this process, we sent our survey to a total of 623 biology instructors across the United States, representing a total of 326 institutions (142 research institutions, 102 associate degree granting colleges, and 83 minority serving institutions). In addition to these efforts, we emailed the listserv of the Society for the Advancement of Biology Education Research (SABER), an international organization that supports biology education research, requesting participation from all active members. Using snowball sampling techniques (Etikan et al., 2016), we asked participants for recommendations of other individuals who teach biology to undergraduates in the United States. We received a total of 103 responses to our survey over a period of two weeks in the spring of 2021.

Data Collection

Data collection and research was approved by Auburn University's Institutional Review Board #21-074 EX 2102. The survey we used for this study was developed and piloted by a team of researchers within the Equity and Diversity in Undergraduate STEM (EDU-STEM) Research Coordination Network (referred to as EDU-STEM; Authors TL, EPD, RY, AE, SC, CC, AGD, SF, KSH, AKL, EL, SJM, ST, & CJB). Prior to broadly distributing the final version of the survey, we piloted the survey to current members of EDU-STEM to request feedback on how we could make the survey clearer and to help ensure the questions we asked were being interpreted as intended. We lightly modified some language in response to their comments.

We first asked participants a series of demographic questions, followed by questions concerning their institution types and teaching experience (Table 1). Next, we inferred which assessments instructors changed during the emergency transition to remote instruction, and how and why they changed them. We then asked, for each type of assessment, whether they would keep the changes in subsequent semesters. Finally, participants responded if they made changes with the concern of equity in mind, and if so, to explain their equity-related concerns. We deliberately placed the question about equity and assessments at the end of the survey as to avoid inadvertently priming instructors to consider equity when responding to the other questions. Data from participants who only responded to some, but not all, of the questions were still analyzed for our study (for the full survey, see the supplemental materials).

Analyses

(1). How did assessment change during the emergency transition to online learning due to COVID-19?

To address this research question, we asked three survey questions (supplemental materials). First, we asked “BEFORE COVID-19 how did you assess students? (select all that apply)”. We then followed this question by asking “AFTER the emergency transition to remote instruction, check all of the following that you used to assess students. (select all that apply)”. The selection options for both questions were exams, quizzes, assignments, participation, projects, and other. We then used the selections from individual instructors for before and after the emergency transition to remote instruction to report whether they removed, added, or did not change that selected assessment component. Lastly, we asked instructors, “Which of the following did you change in response to the emergency transition to remote instruction caused by COVID-19 in Spring 2020? (select all that apply)”. This survey question allowed instructors to select exams, quizzes, assignments, textbook, participation, projects, or “other” with the option to fill in a blank box. Instructors could select all forms of assessment that applied. We decided that the meaning of terms such as “exams” and “quizzes” were self-evident, so we did not include examples for these assessment forms. We did include examples for other selections in the survey to aid instructors in categorizing forms of assessment: assignments (e.g., papers, homework), textbook (or content resource), participation (e.g., attendance, personal response systems), and projects (e.g., presentations). Each form of assessment instructors selected was followed by two additional questions. For example, if an instructor participant selected “exams”, they would be asked: (1) “You indicated that you changed exams as part of your assessment of students in Spring 2020. HOW did you change exams for assessment? Please explain.”; (2) “WHY did you change exams for assessment? Please explain.”

Three of the authors (TL, RY, and AE) individually reviewed all the instructors' responses to the open-ended “HOW” question for exams and participation selection and generated codes using inductive coding (Creswell, 1994). These two selections had the largest number of responses for analysis out of the previously listed selection options (together, 61% of instructors selected one or both forms of assessment). The researchers then convened to compare codes and develop one unified coding rubric. Using the unified rubric, TL, RY, and AE coded 15-30 responses or “blocks” individually. They then met together to compare their codes and revise the rubric. The researchers used constant comparison methods to ensure quotes within a code were not too different from each other to warrant the creation of a new code (Glesne & Peshkin, 1992). This process was repeated until the group was confident with their rubric. At that point, the three authors coded all the responses to each question separately, then compared all coded answers and coded to consensus. Prior to coding to 100% consensus, percentage agreement increased over time, growing from 65% to 90% for exams and 68% to 85% for participation. In qualitative analyses of how instructors changed exams during the emergency transition to remote instruction, eight codes emerged (Figure 2). For how instructors changed participation during the emergency transition to remote instruction, five codes emerged through analysis (Figure 3).

(2). Were any assessment changes motivated by equity concerns?

Instructors responded to the survey question, "Were any of these changes you made to assessment motivated by equity concerns? (exams, quizzes, assignments, textbook/content resource, participation, presentations) Please explain." This survey question was an open-ended response, and participants could respond with as much text as needed. First, we categorized instructor responses based on whether the changes they made were motivated by equity concerns (yes) or were not (no). If instructor changes to assessment were motivated by equity concerns, two of the authors (TL and RY) individually reviewed all these responses and generated codes using inductive coding (Creswell, 1994). The researchers then convened to compare codes and develop one unified coding rubric. Using the unified rubric, TL and RY coded individually. They then met together to compare their codes and revise the rubric. The researchers 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 the group was confident with their rubric. At that point, the two authors coded all the responses to each question separately, then compared all coded answers and coded to consensus. Prior to coding to 100% consensus, percentage agreement was 87%. Six codes emerged (Figure 4).

(3). How did potential changes in assessment due to COVID-19 impact instructor choices in subsequent semesters?

Instructors responded to the following two open-ended survey questions, “If you are going to teach in the remote instruction format in the future, please explain which changes you will keep and which ones you will not keep. Why?” and “If you are going to teach in a face-to-face format in the future, please explain which changes you will keep and which ones you will not keep. Why?” We observed that instructors generally responded to one or both questions by explaining what they would *keep* in remote instruction format or face-to-face format. To understand what forms of assessment instructors kept in the future for remote instruction and face-to-face formats, two authors, (TL and AE) individually reviewed all these responses and generated codes using inductive coding (Creswell, 1994). The researchers then convened to compare codes and develop one unified coding rubric. Using the unified rubric, TL and AE coded individually. They then met together to compare their codes and revise the rubric. The researchers 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 the group was confident with their rubric. At that point, the two authors coded all the responses to each question separately, then compared all coded answers and coded to consensus. Percentage agreement increased over time, growing from 67% to 83% until finally reaching consensus. Five codes emerged representing different assessment forms (Figure 5).

Finally, we asked instructors in the survey to upload a copy of two syllabi: one from a ‘traditional’ semester before the pandemic (i.e., Fall 2019 or earlier) and one from a semester after the emergency transition to remote instruction due to COVID-19 (i.e., Fall 2020 or later) .

We obtained 33 pairs of syllabi which we analyzed for (1) the presence of diversity statements and (2) the change in syllabi grading rubrics for exams, quizzes, participation, assignments, and projects in percentage of final grade between pre and post COVID-19 syllabi. Using the grading rubrics in each syllabus, we sorted the different forms of assessment into the five categories of assessment we listed previously (Figure 5). We also recognize that there could be many opportunities for instructors to implement inclusion statements in different parts of the syllabi. However, we looked explicitly for sections labeled as diversity statements in syllabi for this study.

Results

(1). How did assessment change during the emergency transition to online learning due to COVID-19?

We found that overall, instructors were more likely to report removing forms of assessment rather than adding them. Out of the 70 respondents that responded to this specific question, 46 indicated they removed course components and 20 indicated they added components (Figure 6A). Instructors most commonly removed participation as a form of assessment, with 20 instructors reporting removing it (Figure 6A). The only type of assessment that more instructors reported adding than removing after the emergency transition to remote instruction was assignments (4 additions and 3 removals; Figure 6A). We observed no reports of instructors *adding* exams as a form of assessment.

Because instructors made the most assessment changes to exams and participation, we wanted to better understand these changes. We categorized instructor responses to open-ended questions that asked *how* instructors changed exams and participation. A large percentage of instructors transitioned to online exams (61.54%), and specifically we saw a transition to open-note exams (33.33%). For exams, we also saw a reported decrease in proportion of final grade (7.69%; Figure 6B). For participation, we saw a large addition of virtual participation (39.47%). Instructors also decreased the amount that participation accounted for in students' final grades (28.95%; Figure 6C).

(2). Were any assessment changes during the emergency transition motivated by equity concerns?

For our second research question, we analyzed instructor responses and found that 81.49% of instructors stated that their changes to assessment were motivated by equity concerns (Figure 7). Through further analysis, we thematically coded the "yes" responses to find that the highest concern for equity was access to learning resources (47.73%). Work (13.64%) and family care (11.36%) were other reported equity concerns (Figure 7).

(3). How did potential changes in assessment due to COVID-19 impact instructor choices in subsequent semesters?

We found that for all the responses to the face-to-face future instruction question, instructors responded that they would keep the changes they made to assignments (12.96%), exams (25.90%), participation (29.63%), projects (3.70%), and quizzes (11.11%); (Figure 8). For all the responses to the online future instruction question, instructors responded that they would keep the changes they made during the emergency transition to remote instruction to assignments (21.57%), exams (27.40%), participation (25.94%), projects (7.84%), and quizzes (17.65%); (Figure 8).

We also saw a dramatic increase in percentage of final grade for specifically assignments from pre to post COVID-19 syllabi. Instructors increased the amount that assignments accounted for in the final grade by 10%. We also saw similar decreases in percentage of final grade for exams and quizzes from pre to post COVID-19 syllabi. Instructors decreased the amount that exams accounted for in the final grade by 5% and quizzes in the final grade by 4%. For diversity statements of all submitted syllabi, we saw that 31% had statements in both pre- and post-COVID-19 syllabi, 63% did not have statements for either pre- or post-COVID-19 syllabi, and only 6% changed to having one (Figure 9). Unfortunately, only 11 of our 33 pairs of syllabi had any presence of a diversity statement. For this reason, we felt our sample size was too small to analyze any presence or absence of diversity statements.

Discussion

The COVID-19 pandemic occurred mid-semester, forcing instructors to navigate multiple waves of challenges immediately after the closure of institutions through the end of the semester, including how to adapt their grading structure to a remote-learning environment. This crisis in higher education led us to investigate whether instructors changed how they assessed students during the emergency transition to online learning and if these potential changes were motivated by equity concerns. Specifically, we asked the following three questions: **(1)** How did assessment change during the emergency transition to online learning due to COVID-19? **(2)** Were any assessment changes during the emergency transition motivated by equity concerns? **(3)** How did potential changes in assessment due to COVID-19 impact instructor choices in subsequent semesters? We explored these questions through a survey of undergraduate biology instructors from a variety of institution types across the United States. Below we describe the most prominent themes from survey results.

Instructors often removed participation from the grading structure

Our surveys revealed that instructors were more likely to remove components of assessment than add them. They most commonly removed participation as a form of student assessment, which is a pillar of active learning pedagogy (Driessen et al., 2020a). By eliminating participation requirements, students had little incentive to attend class and contribute meaningfully (Wyatt, 2021), though we recognize that instructors may have included non-graded forms of participation. Previous work shows increased participation and active learning are beneficial for all students (Freeman et al., 2014), particularly those who identify with groups historically marginalized in or excluded from science (Ballen, Salehi & Cotner, 2017; Haak et al., 2011; Theobald et al., 2020). However, the emergency transition to online learning left instructors with limited time to adapt their courses to a new format (Basilaia et al., 2020; Dhawan, 2020; Johnson et al., 2020). The removal of graded participation opportunities could be due to course format changes such as the switch from in-person instruction to asynchronous online instruction. The decision to conduct the course in an asynchronous format may have been intended to increase safety and access for students who would not have otherwise been able to attend classes due to schedule and time zone changes. However, the asynchronous format meant that students lost opportunities to work and study with each other in class (Driessen et al., 2020b) and there were fewer opportunities for students to display their knowledge for the purposes of evaluation and peer discussion (Dumford & Miller, 2018). So, while instructors may have had students' best interests in mind when they eased or eliminated participation requirements, student learning overall may have suffered.

Instructors changed exam format to open-note

Aside from removing graded participation opportunities, instructors commonly changed the exam format from closed-note to open-note. Here we consider open-note exams as those where students can consult textbooks, notes, or other course–related material during the exam. Open-note tests focus on targeting and developing higher-level skills such as conceptualization, problem solving, and reasoning while closed- note exams usually do not emphasize these levels of learning (Feller, 1994). Advocates of open-note exams suggest it rewards the ability to gather and critically analyze material from multiple sources, as opposed to close-note exams which positively recognize short-term storage and rapid retrieval (Ambrose et al., 2010; Krasne et al., 2006; Theophilides & Koutselini, 2000). Open-note exams decrease students' exam anxiety (Block, 2011; Gharib et al., 2012; Williams and Wong, 2007) and minimize a desire to cheat (King, Guyette & Piotrowski, 2009; Nguyen et al., 2020; Watson & Sottile, 2010). While some work shows open-note exams increase performance (Agarwal et al., 2008; Agarwal & Roediger, 2011), others suggest diminished long-term retention of material (Moore and Jensen, 2007). Overall, previous research has mixed claims about open-note exams (Block, 2012; Eilertsen & Valdermo, 2000; Williams & Wong 2007).

Instructors lowered the stakes on exams

Another common example of how instructors changed exams was increased time for students and multiple attempts (Chadha et al., 2020; Williams and Wong, 2007). Fewer high stakes forms of assessment such as exams and quizzes can decrease anxiety and alleviate gender disadvantages for students (Ballen, Salehi & Cotner, 2017; Salehi et al., 2019). Instructors also decreased the proportion that exams accounted for in students' final grades (Bancroft et al., 2019; von der Embse et al., 2015) and increased the number of critical thinking questions (Momson et al., 2010). Previous research shows decreasing reliance on high stakes exams and implementing mixed methods of assessment benefits students (Cotner & Ballen, 2017; Haak et al., 2011; Lyons et al., 2022). These decisions increased flexibility of assessments in order to help address challenges students faced during the pandemic such as access to adequate technology (Van Dijk 2006), computer anxiety (Brosnan, 1998), and mental health issues (Lischer et al., 2021).

Instructors were often concerned with equitable student access to learning resources

Most instructors reported that changes to assessment were made with equity concerns for their students in mind. However, only so many changes could be made during the emergency transition, which took place over days or weeks. The most frequently reported equity concern that instructors had for students was access to learning resources. As reported in previous studies, students reported difficulties from loss of access to learning resources such as Supplemental Instruction sessions, adequate internet access, and library access (Driessen et al., 2020b). Although these were major concerns, instructors were not able to account for these problems because of the mandatory nature of the transition online.

Lessons from a pandemic: lasting changes in biology classrooms

As researchers recognize the need to better understand how change occurs in STEM higher education, we demonstrate dramatic change due to a global pandemic. Interestingly, our results show how instructors intend to keep several changes they made during this chaotic period, and many of those changes are equitable and evidence-based. Instructors report a continued decrease in the amount exams accounts for the final grade, a continued increase in the amount assignments accounts for the final grade, and continued addition of virtual participation. With this study, we hope to add to the body of research that documents what changed, why those changes were made, and what course modifications instructors intend to maintain in future semesters in response to this unique rapid transitional period. In this way, we can be prepared for another factor of immediate pedagogical change and assess students in the most equitable and evidence-based way possible.

Limitations

We acknowledge several limitations in this study. First, because some states do not have institutions that fit into our pre-assigned categories, and because some institutions do not have up-to-date websites with contact information of current faculty, we were not able to obtain all the data we originally set out to collect. Second, we attempted to distinguish between assessment types in our survey by providing examples for each type, but we understand that some instructors could place unique assessment practices into one category while others place them into a different category. Third, for pre and post COVID-19 syllabi, we realize our sample represented a subset of the total dataset (i.e., N=33 syllabi and N=103 survey respondents). Therefore, our findings may not be representative of all the participants. Finally, some of the questions in our survey investigate assessment practices after an institution's transition to remote instruction. However, some states and institutions transitioned online earlier than others based on several reasons, which could affect instructors' experiences of this time period and their interpretation of our survey questions.

Conclusions

We found instructors changed several of their assessment strategies because of the emergency transition to remote instruction due to COVID-19. However, instructors often reported their assessment choices were motivated by equity concerns for their students. In some cases, the motivation to be more equitable may not align with the evidentiary basis (e.g., regarding student participation), suggesting a potential avenue for future research and communication.

Rather than focusing on instructor responses to professional development or other support systems, this work provides important insights into the real-life decision-making processes that instructors face on a regular basis. In times of rapid change, instructors are often forced to implement changes without extensive support or the time needed to reflect on the literature-based in a given area. By focusing on instructional choices during real-world situations, we can better understand how to support our colleagues, in the best interests of our students, during COVID-19 and in future disruptions.

How would you describe your institution type?	
	Community College (25.35%)
	Master's-granting Institution (5.63%)
	PhD-granting Institution (39.44%)
	Primarily Undergraduate Institution (28.17%)
	Other (1.41%)
Is your institution a Minority Serving Institution?	
	Yes (42.86%)
	No (57.14%)
What is your current position?	
	Faculty (90.67%)
	Postdoc (4.00%)
	Graduate Student (2.67%)
	Other (2.67%)
What type of biology classes do you teach? (select all that apply)	
	Graduate level (38.03%)
	Upper level (requires at least one prerequisite; 9.86%)
	Lower level (introductory; 36.62%)
	Multiple (15.49%)
What size biology classes do you teach? (If you teach more than one class, please apply the rest of the survey to the most introductory class)	
	Small (50 or fewer students) (56.94%)
	Medium (more than 50 up to 100 students) (8.33%)
	Large (more than 100 students) (18.06%)
	Multiple (16.67%)

Table 1. Information about biology instructors who participated in the survey.

Figure 1. We used a novel sampling method to recruit instructors by systematically sampling from different institution categories. The dots are color coded for each institution type and in each state that is represented. We also incorporated snowball recruitment. We developed this approach for the purpose of increased representation of instructors who teach undergraduate biology students across the United States.

Codes	Codes Explanation	Instructor Excerpt Example
Decreased Essay/Short Answer	References either changing their exams to have more multiple choice/fill in the blank/simple testing strategies, or references having fewer essay/short answer/complex testing strategies	"Strictly online with fewer essay questions - more multiple choice, matching, fill-in-the blank"
Decreased Multiple Choice	References either changing their exams to have more essay/short answer/complex testing strategies or references having fewer multiple choice/fill in the blank/simple testing strategies	"No longer proctored. More critical thinking questions, more open ended questions"
Decreased or Removed Exams	References decreasing number of exams or moving to zero exams altogether	"Prior to covid, I gave in class multiple choice. Post covid in online learning, NO exams"
Transition to Open-Note	References going closed book to open book or closed notes to open notes. Essentially less restrictions on exams	"When we went online in Spring we transitioned from closed book proctored timed exam to openbook un-proctored and un-timed exams"
Increased Critical Thinking	Increased the critical thinking being used by students taking their exams. Usually done by making more complex questions and trying to adjust for cheating	"No longer proctored. More critical thinking questions, more open ended questions"
Transition to Online	Simply stated that one change was the switch to online format for taking exams. This can be the only thing they say or along with things they changed	"They all switched to online with electronic proctoring from paper exams with student drawings in the past"
Time and Attempt Flexibility	References increased time or more variable amounts of time allotted for exams or more attempts to possibly take exams for students	"They are all online through the Canvas software, I also allowed for students to take the exam all day"
Decreased Proportion of Final Grade	Mentions that the weight of exams in the rubric was decreased	"1. students get multiple days instead one class period, 2. students get 2 attempts rather than 1, 3. the exams pull questions from multiple test banks instead of one version, 4. weight of score was less in final grade rubric"

Figure 2. Instructors' reports of how they changed exams after the onset of the COVID-19.

pandemic. We include explanations of each code, along with an example. Note, color codes of each thematic code correlate with colors in Figure 6B.

Codes	Codes Explanation	Instructor Excerpt Example
Addition of Virtual Participation	References adding new sources of participation such as zoom, virtual polling, discussion boards, etc.	"In the past I've used in-class questions a lot. Now I use play posit videos, discussion boards, and surveys"
Attendance Taken Virtually	References attendance being mandatory on Zoom	"The modified participation included attendance for Zoom calls"
Decreased Attendance Expectations	References lowering attendance requirements or removing altogether	"In class attendance points were dropped when we moved to Zoom, and there is no expectation of students coming to Zoom"
Extra Credit Participation	References making new extra credit opportunities in regards to participation	"Students that were active in class (Zoom cameras on, answered questions, etc.) were awarded an exam grade when attendance dates were added up. Points were assigned for every class lecture so those points were used as an exam score"
Decreased Proportion of Final Grade	References lowering participation requirements, removing altogether, or lowering weight of participation grades	"We went asynchronous. No participation component"

Figure 3. Instructors' reports of how they changed student participation after the onset of the COVID-19 pandemic. We include explanations of each code, along with an example. Note, color codes of each thematic code correlate with colors in Figure 6C.

Code	Codes Explanation	Instructor Excerpt Example
PEERs	States concern for PEER students led to changes	"One goal has been to increase persistence of URM students in STEM by providing a better support network in the class room and changing the evaluation mechanisms"
Health and Anxiety	Response states that concern about student health/anxiety led to changes in assessment	"I wanted to reduce anxiety for students dealing with trauma. Honestly we are all dealing with trauma now, but it's helped make me more aware of trauma informed practices that I plan to carry with me into all my future teaching"
Family Care	States that changes were motivated by concerns for student's family concerns	"Some were living in enviroments where they had to take on roles as big brothers/sisters, mothers, care takers, etc. COVID-19 has caused major shifts in our and students family situations"
Work	States that concern for student employment affected their change	"Many were done so students could work and take courses at the same time. Most of my students really needed the extra income to support their family members, many of whom lost their jobs due to the pandemic"
No Explanation	Response says yes but does not elaborate to fit other codes	"Nearly all changes had equity as a concern. I'm teaching a hybrid class with some students face to face and other fully remote"
Access to Learning Resources	States that changes were related to concern about access to course material	"Most of these were motivated by equity and access concerns"

Figure 4. Types of equity concerns reported by instructors as they made changes to assessments after the onset of the COVID-19 pandemic. We include explanations of each code, along with an example. Note, color codes of each thematic code correlate with colors in Figure 7.

Assessment Codes	Instructor Excerpt Example	
	Face-to-Face	Remote
Assignments (e.g., papers, homework)	"I will continue with the elimination of paper based work. All homework and papers will be uploaded or completed within the LMS"	"I do plan on keeping the online lab manual and online assignments, as they can be used, at the least, as supplemental practice for my students"
Exams	"Next Fall (2021) if we are face-to-face I think we will likely go back to the traditional timed closed book exams. This is so that students grades are comparable to previous years. However I think we do need to rethink this because of student anxiety around exams"	"I will keep the assessment online because it gives students a window of opportunity to take the test"
Participation (e.g., attendance, personal response systems)	"If students are allowed to Zoom for the lecture, I will most likely keep the required camera/mic on (as needed) for attendance"	"I will keep the same attendance policy but expect more participation in class. At this point, I think a lot of students are better equipped for online instruction and can handle having their cameras on"
Quizzes	"I like the changes that I made to the quizzing format. I think it's better for the students' learning, so I will keep that"	"I'd like to shift more towards quizzes and away from exams"
Projects (e.g., presentations)	"The student presentations went much better online, and student gave more feedback and asked more questions by chat and voice. They were very comfortable with this format. I would also keep at least one face to face presentation, because it's clear they are less comfortable and should practice for professional development"	"Will keep virtual lab activities/projects so that students at least learn something"

Figure 5. Changes in assessments that instructors kept in face-to-face or remote learning formats after the COVID-19 semester, as well as excerpts from open responses. Note, color codes of each assessment code correlate with colors in Figure 6A.

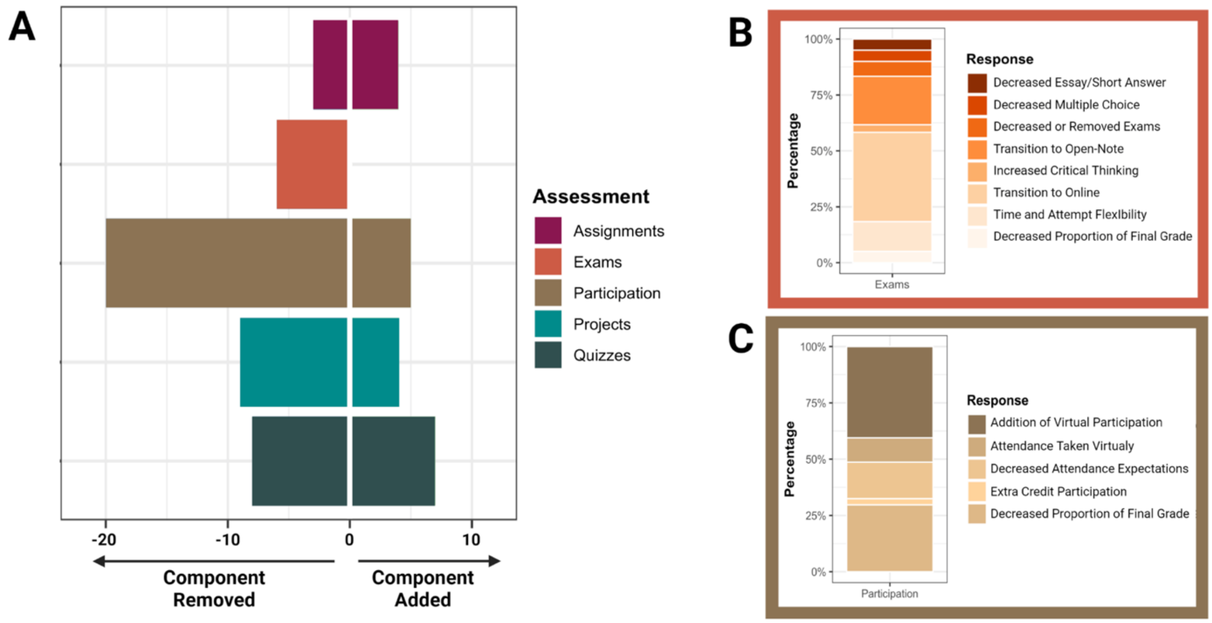

Figure 6. Assessment components removed, added, and modified during the transition to online learning due to COVID-19. (A) Number of times instructors removed or added assessment components: assignments (4 added, 3 removed), exams (0 added, 6 removed), participation (5 added, 20 removed), projects (4 added, 9 removed), and quizzes (7 added, 8 removed). Instructors also maintained (i.e., neither removed or added) assessment components but changed them as they adapted their courses to an online format. We coded open-ended responses that described changes to exams and participation in more detail because they were the most common types of assessments that instructors changed due to the constraints of the pandemic. (B) Coded responses for how instructors changed exams, sorted into eight categories; (C) Coded responses for how instructors changed participation, collapsed into five categories. Note, the order of the parallel bar legends corresponds with the order of the bars in the graph.

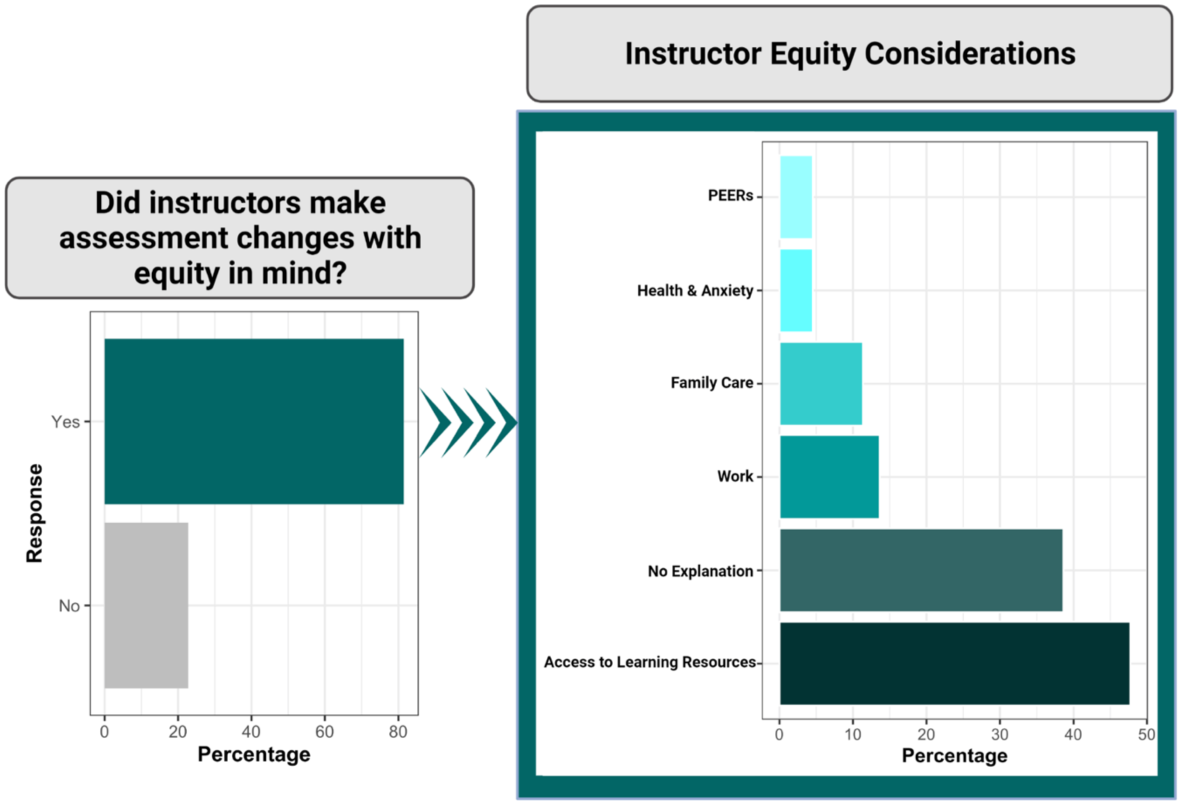

Figure 7. Percentage of instructor responses collapsed into either yes or no categories for changes made with equity in mind. Percentage of what instructors explained their equity concerns to be when responding with yes collapsed into six categories: PEERs, Health & Anxiety. Family Care, Work, No Explanation, and Access to Learning Resources.

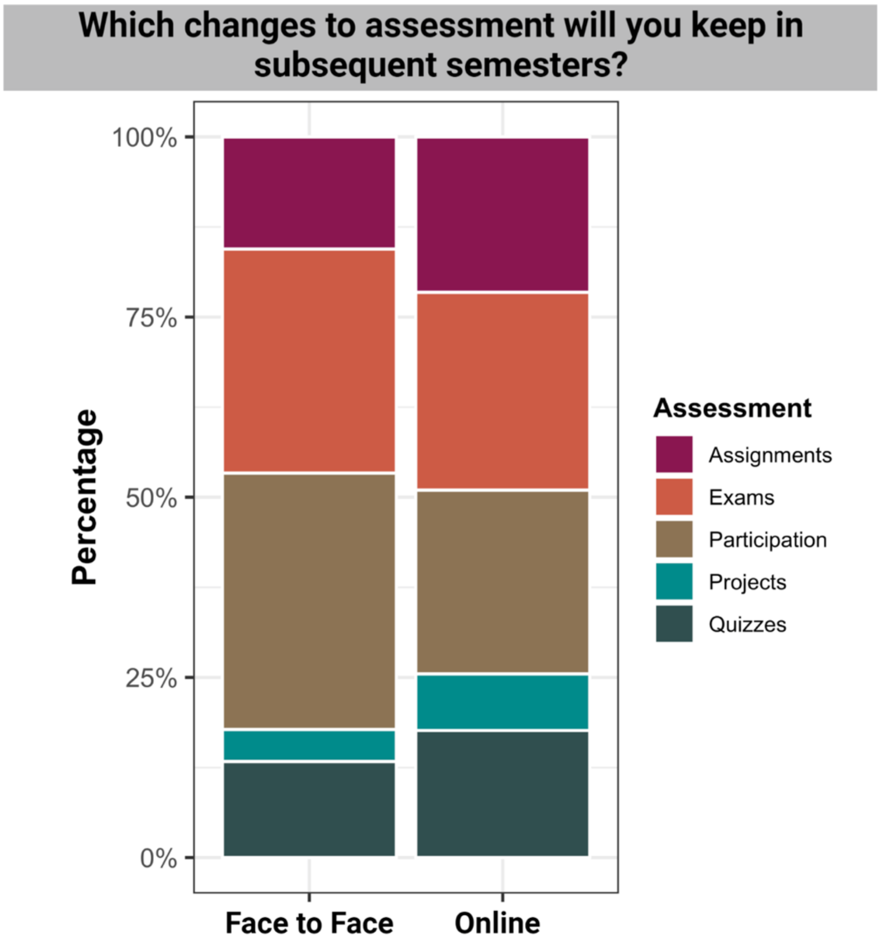

Figure 8. Changes to assessment that instructors intend to keep in subsequent semesters either in face-to-face classes or in online classes. We collapsed percentages of instructor responses into five overarching categories of assessment types: Assignments, Exams, Participation, Projects, and Quizzes.

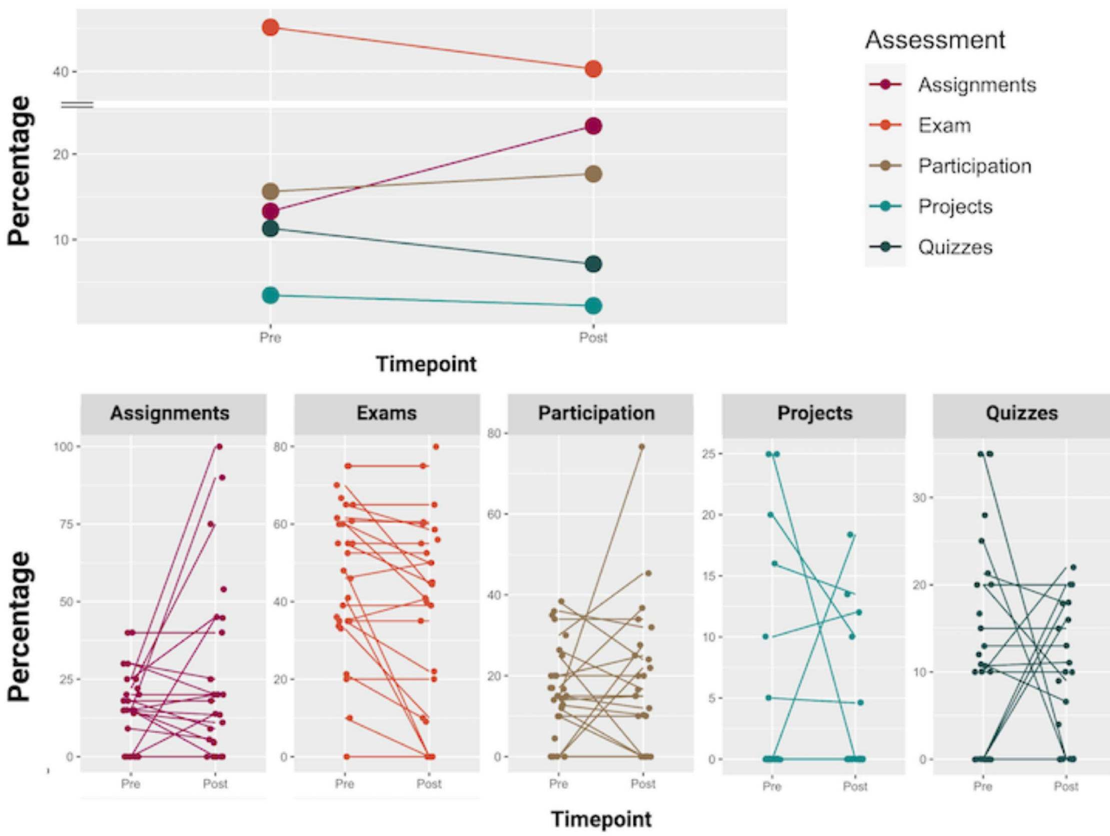

Figure 9. Instructor grading rubric changes for Assignments, Exams, Participation, Projects, and Quizzes in paired instructor syllabi from before COVID-19 (pre) and after COVID-19 (post). Percentages show the amount that each assessment accounts for in students' final grades. Top panel: the average change across all syllabi before and after COVID-19; Lower panels: values from individual syllabi.

Works Cited

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger III, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 22*(7), 861-876.

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

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, 2022.

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

Aucejo, E. M., French, J., Araya, M. P. U., & Zafar, B. (2020). The impact of COVID-19 on student experiences and expectations: Evidence from a survey. *Journal of public economics*, *191*, 104271.

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS ONE*, 12(10), e0186419. https://doi.org/10.1371/journal.pone.0186419

Bancroft, S. F., Fowler, S. R., Jalaeian, M., & Patterson, K. (2019). Leveling the field: flipped instruction as a tool for promoting equity in general chemistry. *Journal of Chemical Education*, *97*(1), 36-47.

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), ar38.

Basilaia, G., Dgebuadze, M., Kantaria, M., & Chokhonelidze, G. (2020). Replacing the classic learning form at universities as an immediate response to the COVID-19 virus infection in Georgia. *International Journal for Research in Applied Science and Engineering Technology*, *8*(3), 101-108.

Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., ... & Risley, 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.

Block, R. M. (2012). A discussion of the effect of open-book and closed-book exams on student achievement in an introductory statistics course. *Primus*, 22(3), 228-238.

Brosnan, M. J. (1998). The impact of computer anxiety and self-efficacy upon performance. *Journal of computer assisted learning*, *14*(3), 223-234.

Carnegie Classification of Institutions of Higher Education. (n.d.). *About Carnegie Classification*. Retrieved March 20, 2021, from http://carnegieclassifications .iu.edu/

Chadha, D., Maraj, M., & Kogelbauer, A. (2020). Opening up Assessment in the Age of COVID-19: Exploring the Utility of Online Open-Book Exams. *Advances in Engineering Education*, *8*(4), n4.

Chittooran, M. M. (2020). Education for Equity: Implications for Teacher Education Programs in the United States. Journal of Academic Perspectives Volume, (3), 1.

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable?. *PLoS One*, *12*(12), e0189610.

Creswell, J. W., & Creswell, J. D. (1994). Research design.

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5-22.

Driessen, E., Beatty, A., Stokes, A., Wood, S., & Ballen, C. (2020a). Learning principles of evolution during a crisis: An exploratory analysis of student barriers one week and one month into the COVID-19 pandemic. *Ecology and evolution*, *10*(22), 12431-12436.

Driessen, E., Knight, J. K., Smith, M. K., & Ballen, C. J. (2020b). Demystifying the Meaning of Active Learning in Postsecondary Biology Education. *CBE Life Sciences Education*, 19(4), ar52. https://doi.org/10.1187/cbe.20-04-0068

Dumford, A. D., & Miller, A. L. (2018). Online learning in higher education: exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education*, *30*(3), 452-465.

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

Etikan, I., Alkassim, R., & Abubakar, S. (2016). Comparision of snowball sampling and sequential sampling technique. *Biometrics and Biostatistics International Journal, 3*(1), 55.

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

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*, *111*(23), 8410. https://doi.org/10.1073/pnas.1319030111

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.

Gharib, A., & Phillips, W. (2012). Test anxiety and performance on open book and cheat sheet exams in introductory psychology. *IPEDR*, 53(1), 1-4.

Glesne, C. Peshkin. (1992). Becoming qualitative researchers: An introduction. White Plains.

Goubeaud, K. (2010). How is science learning assessed at the postsecondary level? Assessment and grading practices in college biology, chemistry and physics. *Journal of Science Education and Technology*, *19*(3), 237-245.

Gurung, R. A. R., & Stone, A. M. (2020). You can’t always get what you want and it hurts: Learning during the pandemic. *Scholarship of Teaching and Learning in Psychology.* Advance online publication. https://doi.org/10.1037/stl0000236

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.

Johnson, N., Veletsianos, G., & Seaman, J. (2020). US Faculty and Administrators' Experiences and Approaches in the Early Weeks of the COVID-19 Pandemic. *Online Learning*, *24*(2), 6-21.

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. F., Relan, A., & Drake, T. A. (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Advances in Health Sciences Education,* 11(2), 155-171.

Lashley, M. A., Acevedo, M., Cotner, S., & Lortie, C. J. (2020). How the ecology and evolution of the COVID-19 pandemic changed learning. Ecology and Evolution, 10(22), 12412-12417.

Lassoued, Z., Alhendawi, M., & Bashitialshaaer, R. (2020). An Exploratory Study of the Obstacles for Achieving Quality in Distance Learning during the COVID-19 Pandemic. *Education Sciences*, *10*(9). https://doi.org/10.3390/educsci10090232

Lischer, S., Safi, N., & Dickson, C. (2021). Remote learning and students' mental health during the Covid-19 pandemic: A mixed-method enquiry. *Prospects*, 1-11.

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics, 74*(2), 118-122.

Lyons, E., Mesghina, A., & Richland, L. E. (2022). Complicated Gender Gaps in Mathematics Achievement: Elevated Stakes during Performance as One Explanation. Mind, Brain, and Education.

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., ... & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *Aera Open, 3*(4), 2332858417743754.

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.

Moore, R., & Jensen, P. A. (2007). Do open-book exams impede long-term learning in introductory biology courses?. *Journal of College Science Teaching*, *36*(7), 46.

Nambiar, D. (2020). The impact of online learning during COVID-19: students' and teachers' perspective. *The International Journal of Indian Psychology*, *8*(2), 783-793.

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.

Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020). Minimize online cheating for online assessments during COVID-19 pandemic. *Journal of Chemical Education*, 97(9), 3429-3435.

Patrick, L. E., Howel, L. A., & Wischusen, W. (2016). Perceptions of active learning between faculty and undergraduates: Differing views among departments. *Journal of STEM Education: Innovations and Research*, 17(3).

Pearson, M. I., Castle, S. D., Matz, R. L., Koester, B. P., & Byrd, W. C. (2022). Integrating Critical Approaches into Quantitative STEM Equity Work. *CBE—Life Sciences Education*, *21*(1), es1.

Reinholz, D. L., White, I., & Andrews, T. (2021). Change theory in STEM higher education: A systematic review. *International Journal of STEM Education*, *8*(1), 37. https://doi.org/10.1186/s40594-021-00291-2

Riordan, M., Klein, E. J., & Gaynor, C. (2019). Teaching for Equity and Deeper Learning: How Does Professional Learning Transfer to Teachers' Practice and Influence Students' Experiences?. *Equity & Excellence in Education*, *52*(2-3), 327-345.

Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers. SAGE Publications Inc.* (Vol. XXXIII).

Salehi, S., Cotner, S., Azarin, S. M., Carlson, E. E., Driessen, M., Ferry, V. E., ... & Ballen, C. J. (2019, September). Gender performance gaps across different assessment methods and the underlying mechanisms: The case of incoming preparation and test anxiety. In *Frontiers in Education* (Vol. 4, p. 107). Frontiers.

Schinske, J. N., Balke, V. L., Bangera, M. G., Bonney, K. M., Brownell, S. E., Carter, R. S., Curran-Everett, D., Dolan, E. L., Elliott, S. L., Fletcher, L., Gonzalez, B., Gorga, J. J., Hewlett, J. A., Kiser, S. L., McFarland, J. L., Misra, A., Nenortas, A., Ngeve, S. M., Pape-Lindstrom, P. A., … Corwin, L. A. (2017). Broadening Participation in Biology Education Research: Engaging Community College Students and Faculty. *CBE—Life Sciences Education*, *16*(2), mr1. https://doi.org/10.1187/cbe.16-10-0289

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., 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*, *117*(12), 6476. https://doi.org/10.1073/pnas.1916903117

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open-book examination: 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., Cotner, 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. https://doi.org/10.1187/cbe.20-05-0085

Van Dijk, J. A. (2006). Digital divide research, achievements and shortcomings. *Poetics*, *34*(4-5), 221-235.

von der Embse, N. P., Schultz, B. K., & Draughn, J. D. (2015). Readying students to test: The influence of fear and efficacy appeals on anxiety and test performance. *School Psychology International*, *36*(6), 620-637.

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

Williams, J. B., & Wong, A. (2007). Closed book, invigilated exams versus open book, open web exams: An empirical analysis. *ICT: Providing Choices for Learners and Learning. Proceedings Australasian Society for Computers in Learning in Tertiary Education Singapore*, 1079-1083.

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.

Wyatt, B. (2021). Insights into student participation in a soil physics course during COVID-19 emergency online learning. *Natural Sciences Education*, *50*(1).

Appendix 1. Qualtrics Survey

How would you describe your institution type?
PhD-granting Institution
Primarily Undergraduate Institution
Community College
Master's-granting Institution
Other
Is your institution a Minority Serving Institution?
Yes
No
What is your current position?
Faculty
Postdoc
Graduate Student
Other
What type of biology classes do you teach? (select all that apply)
Graduate level
Upper level (requires at least one prerequisite)
Lower level (introductory)
Other
What size biology classes do you teach? (If you teach more than one class, please apply the rest of the survey to the most introductory class)
Small (50 or fewer students)
Medium (more than 50 up to 100 students)
Large (more than 100 students)
Please upload pre-COVID syllabus (Fall 2019 or before; please remove name/institutional affiliation)

Please upload post-COVID syllabus (Fall 2020 or after; please remove name/institutional affiliation)
To what extent did you change how you assessed students when you transitioned to online teaching? (Please use "1" as no change)
1 (no change)
2
3
4
5
6
7 (dramatic change)
BEFORE COVID-19 how did you assess students? (select all that apply)
Exams
Quizzes
Assignments (e.g., papers, homework)
Participation (e.g., attendance, personal response systems)
Projects (e.g., presentations)
Other
Other
Other
AFTER the emergency transition to remote instruction, check all of the following that you used to assess students. (select all that apply)
Exams
Quizzes
Assignments (e.g., papers, homework)
Participation (e.g., attendance, personal response systems)
Projects (e.g., presentations)

Other
Other
Other
Which of the following did you change in response to the emergency transition to remote instruction caused by COVID-19 in Spring 2020? (select all that apply)
Exams
Quizzes
Assignments (e.g., papers, homework)
Textbook (or content resource)
Participation (e.g., attendance, personal response systems)
Projects (e.g., presentations)
Other
Other
Other
You indicated that you changed [a form of assessment listed above; hereafter x] as part of your assessment of students in Spring 2020. HOW did you change [x]? Please explain.
Why did you change [x] for assessment? Please explain.
If you are going to teach in the remote instruction format in the future, please explain which changes you will keep and which ones you will not keep. WHY?
If you are going to teach in the face-to-face format in the future, please explain which changes you will keep and which ones you will not keep. WHY?
Were any of these changes you made to assessment motivated by equity concerns? (Exams, quizzes, assignments, textbook/content resource, participation, presentations) Please explain.
Thank you so much for your participation! If you are willing to be contacted with additional questions, please provide your email address.

Appendix 2. A call for data-driven networks to address equity in the context of undergraduate biology (In press, formatted for *CBE- Life Sciences Education*).

Contribution: Collected data; Performed analysis; Revised and edited drafts of the paper.

A Call for Data-Driven Networks to Address Equity in the Context of Undergraduate Biology

Seth K. Thompson,[†*] Sadie Hebert,[†] Sara Berk,[‡§] Rebecca Brunelli,[‖] Catherine Creech,[¶] Abby Grace Drake,[#] Sheritta Fagbodun,[@] Marcos E. Garcia-Ojeda, Carrie Hall,[††] Jordan Harshman,[‡‡] Todd Lamb,[‡] Rachael Robnett,[§§] Michèle Shuster,[‖‖] Sehoya Cotner,[†] and Cissy J. Ballen[‡]**

[†]Biology Teaching and Learning, University of Minnesota–Twin Cities, Minneapolis, MN 55455; [‡]Biological Sciences, Auburn University and [‡‡]Chemistry and Biochemistry, Auburn University, Auburn, AL 36849; [§]Biology, University of Washington, Seattle, WA 98195; [‖]Biological Sciences, California State University–Chico, Chico, CA 95929; [¶]Mt. Hood Community College, Gresham, OR 97030; [#]Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853; [@]Biology, Tuskegee University, Tuskegee, AL 36088; **Quantitative and Systems Biology, University of California–Merced, Merced, CA 95343; [††]Biological Sciences, University of New Hampshire, Durham, NH 03824; [§§]Psychology, University of Nevada–Las Vegas, Las Vegas, NV 89154; [‖‖]Biology, New Mexico State University, Las Cruces, NM 88003

ABSTRACT

National efforts to improve equitable teaching practices in biology education have led to an increase in research on the barriers to student participation and performance, as well as solutions for overcoming these barriers. Fewer studies have examined the extent to which the resulting data trends and effective strategies are generalizable across multiple contexts or are specific to individual classrooms, institutions, or geographic regions. To address gaps in our understanding, as well as to establish baseline information about students across contexts, a working group associated with a research coordination network (Equity and Diversity in Undergraduate STEM, EDU-STEM) convened in Las Vegas, Nevada, in November of 2019. We addressed the following objectives: 1) characterize the present state of equity and diversity in undergraduate biology education research; 2) address the value of a network of educators focused on science, technology, engineering, and mathematics equity; 3) summarize the status of data collection and results; 4) identify and prioritize questions and interventions for future collaboration; and 5) construct a recruitment plan that will further the efforts of the EDU-STEM research coordination network. The report that follows is a summary of the conclusions and future directions from our discussion.

INTRODUCTION

Science teaching in higher education faces many challenges, from inequitable student access to the social polarization of science (Gross, 2006; Mervis, 2011). In the context of undergraduate biology, these challenges are magnified by persistent gaps in performance and degree attainment among members of historically underrepresented groups (Trapani and Hale, 2019). Classroom challenges and institutional barriers impact members of underrepresented groups disproportionately and contribute to observed disparities in higher education (Allen, 1992; DesJardins *et al.*, 2002). Investigations of these barriers for students should be expanded beyond the traditional venue of research-intensive institutions to include other learning environments that serve the large undergraduate population in the United States (Schinske *et al.*, 2017). While a number of committed efforts show promise in promoting historically underserved groups (Wilson *et al.*, 2012; Hernandez *et al.*, 2013; Snyder *et al.*, 2016; Theobald *et al.*, 2020), deliberate evaluation across multiple institutional contexts will

CBE Life Sci Educ December 1, 2020 19:mr2
DOI:10.1187/cbe.20-05-0085

*Address correspondence to: Seth K. Thompson (thom2587@umn.edu).

rigorously assess when change occurs (or not) and inform recommendations for effective evidence-based practices.

To address research priorities, we convened a network of educators and discipline-based education researchers through a research coordination network (RCN) funded by the National Science Foundation (NSF) called Equity and Diversity in Undergraduate STEM (EDU-STEM). EDU-STEM integrates research and teaching in the context of evidence-based classroom experiences across biology curricula. The objectives of EDU-STEM are to: 1) reveal differences, if they exist, in the cultural climate for women and minoritized and marginalized groups in science, technology, engineering, and mathematics (STEM) disciplines (initially focusing on biology) as a function of geography, institution type, and cultural profile of the participating departments; (2) increase the number of faculty in the United States who are familiar with barriers to inclusion in STEM and can apply evidence-based techniques for countering known barriers; 3) develop a community of faculty who can serve as leaders—at their home institutions and nationally—in inclusive teaching and assessment; and 4) identify cultural factors associated with a shift toward evidence-based teaching, especially pertaining to inclusive teaching. In this paper, we present a framework for network activities developed during a meeting of EDU-STEM participants held in Las Vegas, Nevada, in November of 2019.

EDU-STEM MEETING IN LAS VEGAS: INTEGRATING DIVERSE PERSPECTIVES

We convened a meeting of 12 participants to 1) consider the current state of equity and diversity in STEM based on data generated from the incubator year of the grant; 2) reflect on the implications of our results and the value of the network; and 3) decide on future priorities for the network. To maximize our impacts, we invited faculty from community colleges (CCs),[1] institutions with a minority-serving designation (MSIs),[2] and research-intensive institutions (RIs)[3] to attend the meeting. It was important that the network members present reflected the range of institutions integrated into the research network itself. In addition to a number of disciplinary biologists, the group also included discipline-based education researchers and psychologists interested in research on STEM equity and inclusion. Reflecting the diverse and extensive contributions of the network, both meeting participants and other network members are authors of this report. Detailed authorship contributions are provided in the authorship rubric document in the Supplemental Material.

The Equity and Diversity in Undergraduate STEM meeting took place on November 22–23, 2019, and aimed to achieve the following specific objectives:

1. Characterize the present state of equity and diversity in undergraduate biology education research (BER)
2. Address the value of a network of educators focused on STEM equity
3. Summarize the status of data collection and results
4. Identify and prioritize questions and interventions for future collaboration
5. Construct a recruitment plan that will further the goals of EDU-STEM

In the following sections, we describe the results from the meeting related to each of these objectives.

What Is the Present State of Equity and Diversity in Undergraduate Biology Education Research?

Calls for change in education to academically prepare an increasingly diverse student body led to a surge of empirical research on evidence-based teaching (American Association for the Advancement of Science, 2011). Discipline-based education journals such as *CBE—Life Sciences Education*, *Microbiology & Biology Education*, and *CourseSource* and disciplinary biology journals such as *PLoS ONE*, *PLoS Biology*, *BioScience*, *Proceedings of the Royal Society B: Biological Sciences*, and *Science* actively publish BER focused on the undergraduate level.

Without caveats or limitations concerning the population under study, BER assumes students share some fundamental learning processes and that findings from one or a few student populations are applicable across contexts. Critics within other social science fields warn against universal claims about behavioral phenomena when research sampling is based on a single subpopulation, particularly if that pool of participants are from Western, educated, industrialized, rich, and democratic (WEIRD) societies (Henrich *et al.*, 2010a). They argue that psychology (and other social sciences) often make broad statements about fundamental principles of human behavior, when in fact WEIRD populations may be among the most unusual people on Earth (Henrich *et al.*, 2010b). Similarly, if the vast majority of subjects within discipline-based education research are primarily from selective, predominantly white institutions (PWIs), the experiences of students who are white, nondisabled, and middle to upper income will be overrepresented in the literature. Using the experiences of a privileged subset of students as the basis for broad generalizations only further promotes the pervasive dominance of the white experience and unjust power structures in our academic settings. A challenge moving forward for the field will be to test the generalizability of fundamental claims across different student populations.

Thus, one important contribution of EDU-STEM and similar networks is to provide a space to share experiences and develop effective teaching methods from institutions and student populations that are currently (and historically) underrepresented in BER. Furthermore, a network approach allows for the collaborative distribution of resources into institutions and student populations that are most impacted by educational disparities. During the meeting, we questioned the extent to which institutions serving underrepresented minorities (hereafter URMs; which include African-American, American

[1]We define CCs as associate's colleges that offer training and classes that are affordable and relevant to the local community (Schinske *et al.*, 2017; Carnegie Classification of Institutions of Higher Education, n.d.).

[2]MSIs are a federally recognized category of establishment based on minority student enrollment criteria. Examples include historically Black colleges and universities, predominantly Black institutions, Hispanic-serving institutions, Alaska Native and Native Hawaiian–serving institutions, Native American and Pacific Islander–serving institutions, and Tribal colleges and universities (Gasman *et al.* 2008).

[3]For the purposes of the research coordination network, we broadly define RIs as doctoral-granting institutions and master's-granting institutions without a minority-serving designation (Carnegie Classification of Institutions of Higher Education, n.d.).

Indian, Alaska Native, Hispanic/Latinx, underrepresented Asian-American, and other students of color) are represented in/producing the contemporary BER literature. We explored the extent that BER literature focuses on the experiences of students from single institutions, particularly doctoral-granting RIs. We extracted information from recent peer-reviewed literature from 2016, 2017, and 2018 in biology education across three journals ($N = 149$ articles). We collected data from *CourseSource*, a journal that publishes active-learning biology activities for the classroom and laboratories. We also selected two journals that commonly publish BER articles: *CBE—Life Sciences Education* and the *Journal of Microbiology & Biology Education*. To narrow the scope of the inquiry, we focused on studies that addressed some element of active learning in the classroom within undergraduate biology. To do so, we searched for the term "active learning" in the titles, abstracts, or text of research articles and only included papers that focused empirically on pedagogical impacts.

From each article, we collected information such as whether the study focused on a single class or multiple class section(s) or courses (either over time or simultaneously), whether the study took place in one or multiple institution(s), whether the class size was greater or less than 50 students, and whether the focus was on upper- or lower-division classes. We define upper-division classes as those that require a prerequisite. Some papers did not explicitly state the university at which the work was completed; in these situations, we inferred that the study was conducted at the institution where 75% of the authors were working as long as the paper stated in the methods section the geographical range that met the location characteristics of the institution. To characterize institution type, we classified each institution as either a CC, a Baccalaureate Institution, MSI, a doctoral-granting RI, or a master's-granting RI. We used the Carnegie Classification of Institutions of Higher Education website (Carnegie Classification of Institutions of Higher Education, n.d.) to separate research-based institutions into doctoral-granting, master's-granting, or undergraduate universities. Then, we found minority-serving designations through the U.S. Department of the Interior website and reclassified those institutions as MSIs, regardless of their Carnegie Classification.

Our results mirror those from previous studies in BER (e.g., Schinske *et al.*, 2017) and from the psychology literature, showing an overrepresentation of studies from relatively selective RIs (85% of studies); from single classes (62% of studies) within single institutions (93% of studies) that are composed of more than 50 students (77% of studies). Only five studies took place at MSIs (4%), and three studies focused on CC populations (2%; Figure 1). These findings present a critical challenge for the field of discipline-based education research, as the study participants included in most research are not representative of most college students. For example, CCs serve a large proportion of minority, first-generation (FGEN), low-income, and adult students (Ma and Baum, 2016). While MSIs educate 30% of all U.S. undergraduates and produce 20% of the country's STEM bachelor's degrees (National Academies of Sciences, Engineering, and Medicine, 2019), only 4% of all studies took place at MSIs. And while 42% of all undergraduates in the United States are enrolled in CCs (Ma and Baum, 2016), only 2% of BER studies about active learning took place at CCs. Among BER publications that do take place at CCs, Schinske

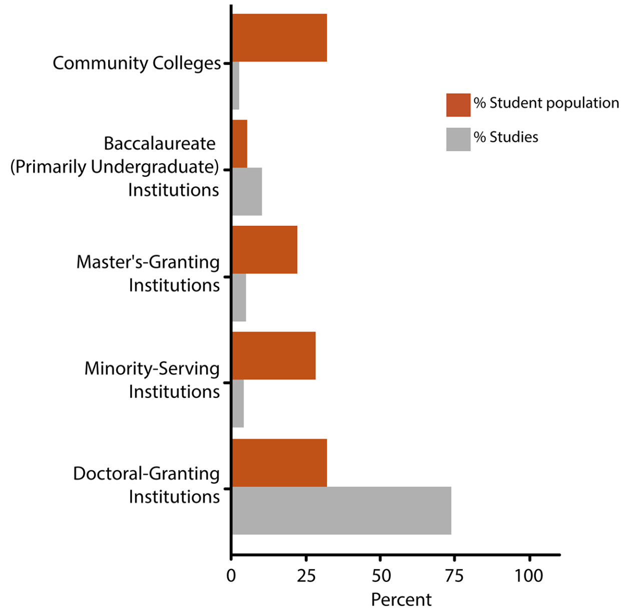

FIGURE 1. The percentage of the student population enrolled in institution types (orange) in the United States and the percentage of contemporary studies about evidence-based teaching that take place at those institutions (gray). Note the overrepresentation of students relative to education research studies at CCs, master's-granting institutions, and MSIs; and the overrepresentation of studies that take place at baccalaureate colleges and doctoral-granting institutions. Note that we pooled all schools that possess a minority-serving designation into one category; all other institutions lack this designation. References: Espinosa *et al.*, 2017; Carnegie Classification of Institutions of Higher Education, n.d.

et al. (2017) reported more than half (51%) of authorship came from individuals not affiliated with CCs. While CCs educate the majority of URM students, research on CC student populations rarely has an explicit focus on equity (Schinske *et al.*, 2017). Such populations may stand to benefit most from research on evidence-based teaching with an emphasis on equitable teaching practices. Given that CCs and MSIs educate the majority of URM students, other institutions can learn from equitable teaching practices that are effective at these institutions. The unintentional mismatch between student populations studied in most BER and student populations enrolled in U.S. institutions of higher education could have important practical consequences as we investigate the largest barriers for students in higher education and develop recommendations for best teaching practices in "typical" college classrooms.

Value of EDU-STEM: Large-Scale Collaboration Focused on Equity and Diversity

EDU-STEM responds to the need to investigate different educational contexts by collaborating with faculty across different institutions and collecting data from a diversity of biology classrooms. As part of EDU-STEM, which has been collecting data from biology classrooms since 2015, we have identified barriers for students and the impacts of evidence-based teaching practices. For example, previous research has found that attrition rates among science majors appear to be highest for members of groups that have a history of underrepresentation in science fields (Seymour and Hewitt, 1994; Chen, 2013). According to recent work, disparities on the basis of race, ethnicity, gender, or FGEN status can be traced to a number of factors related to the classroom climate or the depersonalized, didactic atmosphere that characterizes many undergraduate science courses (Rainey *et al.*, 2018). Learning environments are not only shaped by classrooms but also by institutional cultures, which create conditions wherein groups of people experience unequal opportunities. For example, the extent to which students report judgment from their peers and instructors on the basis of their race might differ based on institutional policies regarding equity and inclusion or the proportion of students who share those identities on campus or the extent that equitable teaching strategies are implemented in classrooms (Massey and Fischer, 2005; Johnson-Ahorlu, 2013). By developing and implementing innovations and reforms informed primarily by research conducted at a single type of institution, we overlook potent forces that likely differ between institution types.

Previous interventions that focus on equity and show promise in one course (e.g., introductory-level, advanced), in one STEM discipline (e.g., physics, biology), and in one setting (e.g., high school, college) may not translate to other instructional contexts. In fact, an intervention that is effective for a specific group of students could be less effective—or may even backfire—for students who have different background attributes (e.g., FGEN vs. continuing generation [CGEN]) or who are situated in a different educational context (e.g., CC vs. RI). In this vein, Steele (1997) cautions against taking a one size fits all approach to academic interventions. Instead, he argues that it is important to tailor interventions to the specific challenges that students encounter.

For example, results from previous studies show certain active-learning strategies reduce or eliminate demographic gaps in performance (Haak *et al.*, 2011; Eddy and Hogan, 2014; Ballen *et al.*, 2017). Until recently, the majority of research addressing this topic has been performed at large RIs and in single courses. Additionally, "active learning" is a broad term and is frequently undefined in the literature. Some active-learning practices increase student anxiety, potentially distracting from learning (England *et al.*, 2017). And without paying explicit attention to equity, active learning can further disadvantage certain underrepresented student groups (Setren *et al.*, 2019; Aguillon *et al.*, 2020). Finally, social aspects of the classroom impact students in different ways, especially those students who tend toward introverted behavior (Beckerson *et al.*, 2020), or who may feel pressure to conceal certain aspects of their identity, such as sexual orientation, political affiliation, or religion (Cooper and Brownell, 2016; Henning *et al.*, 2019).

By leveraging complementary areas of expertise, science educators, psychologists, and data-management specialists can partner to avoid intervention pitfalls at scale. Science educators contribute a deep understanding of science pedagogy and the academic context; psychologists, on the other hand, are trained to identify the ways in which student attributes interact with the educational context to shape academic outcomes, providing insight into psychological mechanisms that account for an intervention's success (or lack thereof); and data-management specialists can help create platforms to assist in the broad interpretation of the results, informing personalized, evidence-based

TABLE 1. Summary of participating institutions included in the EDU-STEM RCN

Institution	Carnegie Classification	MSI
Institution A	Associate's Colleges	Yes
Institution B	Associate's Colleges	No
Institution C	Associate's Colleges	No
Institution D	Master's Colleges and Universities (M1)	Yes
Institution E	Master's Colleges and Universities (M2)	Yes
Institution F	Doctoral Universities (R1)	No
Institution G	Doctoral Universities (R2)	Yes
Institution H	Doctoral Universities (R1)	No
Institution I	Doctoral Universities (R1)	No

teaching practices and their dissemination to other faculty. EDU-STEM originated with a cohort including science educators, psychologists, and data-management experts and intends to grow representation in each of these categories.

Summary of Data Collection and Results

At the 2019 Las Vegas meeting, we discussed data that showed variability across institutions for a variety of student outcomes. The data included student-reported affective characteristics and demographic information collected from surveys and course performance data provided by instructors. The majority of the data was collected during Fall 2017 and Spring 2018 terms, but also included data from the Fall 2015, Spring 2016, Fall 2016, Spring 2017, and Fall 2018 terms. Nine participating institutions contributed information, representing three institution types according to the Basic Carnegie Classification of Institutions of Higher Education along with several designations of MSIs (Table 1).

The analysis (see Methods in the Supplemental Material) included data from 8740 students and confirmed differences across institution types for three student demographic characteristics:[4] gender (female or male), underrepresented minority status (URM or non-URM), and college-going status (FGEN or CGEN). Proportions of female and male students, URM students, and FGEN students varied across institution types (Table 2). For the data collection, students and instructors had the option to omit survey items, which resulted in missing demographic data. Each broad demographic identity was

[4]A note about gender and race categories. For the purposes of this research, we use "male" and "female" to describe gender, recognizing that these refer to biological sex rather than gender. We use these terms, because they more accurately reflect the majority of the data we collected, which were institutional data that often only included binary options. We also use the term "underrepresented minority" to describe students who identify as American Indian/Alaska Native, Black/African American, Latinx/Hispanic/Hispanic American, and Native Hawaiian or Other Pacific Islander. This group excludes Asian/Asian-American and white/European-American individuals. We acknowledge this does not recognize the variation within and among groups. Some individuals in these groups do not identify with this term in a singular way, and some reject this term altogether. Overall a limitation of this research is the nature of these categories, which are problematic, because they are designated by an authority and do not leave room for or recognize people who identify as mixed race or outside the gender binary. Additionally, gender and race are only two of many human social identities that have subpopulations who are minoritized and underrepresented in biology. We plan to address this in the future by encouraging students to self-identify and expanding our categorical descriptors. While imperfect, our categories allow us to establish important baselines of student experiences in biology across institutions.

collapsed for the purposes of this analysis, but we collected fine-grained data on race/ethnicity identity and gender identity, because we realize both of these are complex. Students can identify with multiple racial/ethnic categories, and may identify with more than one gender, or may not identify with a gender.

We investigated differences across institution types for three demographic groups and multiple student outcomes. Here we will focus on two outcomes: survey measures of test anxiety and average exam performance. We measured test anxiety using four items from the Motivated Strategies for Learning Questionnaire (Pintrich, 1991). Across most institutions, females, on average, reported higher test anxiety than males (Figure 2A), and the difference between groups was statistically significant at three institutions (Figure 2, filled circles). URM students, on average, reported higher test anxiety than non-URM students at most institutions (Figure 2B), and the difference between groups was statistically significant at one institution (Figure 2B, filled circles). Interestingly, we found the opposite was true for some institutions. Across all institutions, FGEN students, on average, reported higher test anxiety than CGEN students (Figure 2C), and the difference between groups was statistically significant at one institution. With respect to weighted average exam percent, differences between males and females were highly variable among institutions. Males, on average, outperformed females at six of the institutions; while females, on average, outperformed males at three of the institutions; and the difference between groups was statistically significant at two institutions (Figure 3A). Across all institutions, non-URM students, on average, outperformed URM students, and the difference between groups was statistically significant at four institutions (Figure 3B). Across most institutions, CGEN students, on average, outperformed FGEN students, and the difference between groups was statistically significant at one institution (Figure 3C). In sum, we found evidence for significant performance gaps for multiple populations across multiple institution types, but the differences between groups are highly variable, demonstrating the need for expanded analysis.

Taken together, these results generated a robust discussion about how to interpret the findings and meaningful responses to observed patterns using large-scale collaboration. Following the discussion, several institutional leaders expressed an interest in text anxiety mitigation interventions. Together, this subgroup planned a two-semester exploration that involved: first, identifying whether test anxiety mediates performance, and whether this effect disproportionately impacts historically underserved students, as has been demonstrated by Ballen *et al.* (2017) and Salehi *et al.* (2019); and second, implementing a single intervention, in parallel across their institutions, in hopes of mitigating any demonstrated impacts of test anxiety (e.g., altering the balance of formative versus summative assessment; Cotner and Ballen, 2017). EDU-STEM funds will then allow these subgroup participants to meet to discuss their findings; if warranted, participants can also meet to draft manuscripts and develop next-step plans for classroom interventions.

Research Priorities for Future Collaboration

What research questions are a large-scale collaborative network uniquely positioned to address? After a discussion of the strengths of a research network and the summary of our findings

TABLE 2. Student demographic characteristics by institution

Demographic characteristic	Institution A^ (n = 43)	Institution B (n = 57)	Institution C (n = 467)	Institution D^ (n = 368)	Institution E^ (n = 50)	Institution F (n = 904)	Institution G^ (n = 733)	Institution H (n = 6024)	Institution I (n = 94)
Gender									
Female	33 (76.7%)	23 (40.4%)	248 (53.1%)	233 (63.3%)	31 (62.0%)	278 (30.8%)	322 (43.9%)	2931 (48.7%)	69 (73.4%)
Male	10 (23.3%)	27 (47.4%)	99 (21.2%)	130 (35.3%)	8 (16.0%)	159 (17.6%)	125 (17.1%)	2996 (49.7%)	24 (25.5%)
Missing	0 (0%)	7 (12.3%)	120 (25.7%)	5 (1.4%)	11 (22.0%)	467 (51.7%)	286 (39.0%)	97 (1.6%)	1 (1.1%)
URM status									
Non-URM	16 (37.2%)	45 (78.9%)	245 (52.5%)	216 (58.7%)	0 (0%)	336 (37.2%)	153 (20.9%)	5325 (88.4%)	88 (93.6%)
URM	20 (46.5%)	6 (10.5%)	99 (21.2%)	146 (39.7%)	39 (78.0%)	88 (9.7%)	282 (38.5%)	587 (9.7%)	5 (5.3%)
Missing	7 (16.3%)	6 (10.5%)	123 (26.3%)	6 (1.6%)	11 (22.0%)	480 (53.1%)	298 (40.7%)	112 (1.9%)	1 (1.1%)
First-generation status									
CGEN	0 (0%)	45 (78.9%)	211 (45.2%)	188 (51.1%)	29 (58.0%)	224 (24.8%)	305 (41.6%)	5152 (85.5%)	72 (76.6%)
FGEN	0 (0%)	12 (21.1%)	134 (28.7%)	175 (47.6%)	9 (18.0%)	30 (3.3%)	143 (19.5%)	872 (14.5%)	21 (22.3%)
Missing	43 (100%)	0 (0%)	122 (26.1%)	5 (1.4%)	12 (24.0%)	650 (71.9%)	285 (38.9%)	0 (0%)	1 (1.1%)

^ indicates minority-serving institutions.

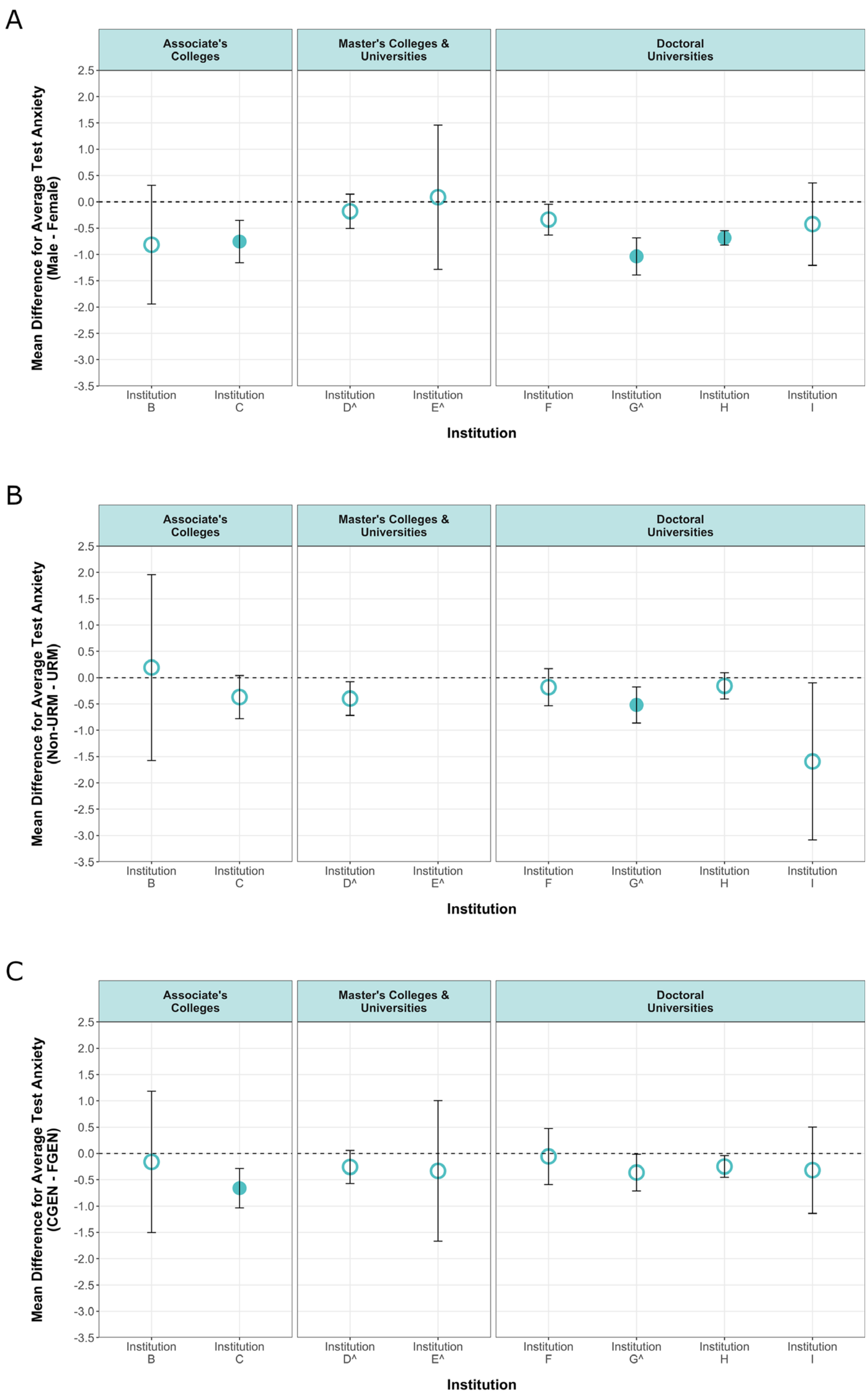

FIGURE 2. Mean differences for average test anxiety across institution type comparing (A) males and females, (B) non-URM and URM students, and (C) CGEN and FGEN students. In each panel, circles represent the differences between group means, and bars represent 95% confidence intervals for the differences between group means. Open circles indicate no significant difference between group means, and filled circles indicate a significant difference between group means. The dotted line represents no difference between groups. Measures below the dotted line indicate (A) females report higher anxiety than males; (B) URM students report higher anxiety than non-URM students; and (C) FGEN students report higher anxiety than CGEN students. MSIs are designated with a ^. A list of deidentified institutions is shown in Table 1.

to date, we identified two research priorities for EDU-STEM moving ahead: 1) context-dependent identity salience, or how salient elements of one's identity impact classroom experiences across different learning contexts; and 2) intersectionality, in which constructs such as race and gender interact with one another *and* with other social categories (e.g., class background) to shape people's experiences in everyday life (Crenshaw, 1989; hooks, 2000). We selected these as worthy research pursuits, because they have direct consequences for student learning and equity, and can be robustly addressed through a collaborative network. We expand on those discussions through a brief literature review of these priorities and develop a case for why collaborative research networks, like EDU-STEM, are poised to address them.

Context-Dependent Identity Salience. Data collected through the EDU-STEM Network provide a unique opportunity to examine how social identities such as race/ethnicity and gender work together to shape students' experiences across institutions of higher education. Research shows that students who identify as URM and women are more likely to experience challenges in STEM higher education settings, but these challenges are documented in largely separate literatures. Research focusing on racial/ethnic disparities indicates that URM students are often underserved in the K–12 education system (Lee and Ransom, 2011; Sáenz and Ponjuan, 2011), which can make for a difficult transition to undergraduate STEM course work. Further, URM students are more likely than white students to be FGEN college students, which can compound other challenges, such as feelings of low belongingness or negative stereotypes about academic ability (Lohfink and Paulsen, 2005). In contrast, research focusing on gender disparities in STEM often focuses on gendered social role expectations. Women inhabit a social system that steers them toward communal careers and roles, which are often perceived as incompatible with STEM achievement (Diekman and Steinberg, 2013). In addition, undergraduate women in STEM contexts report encountering

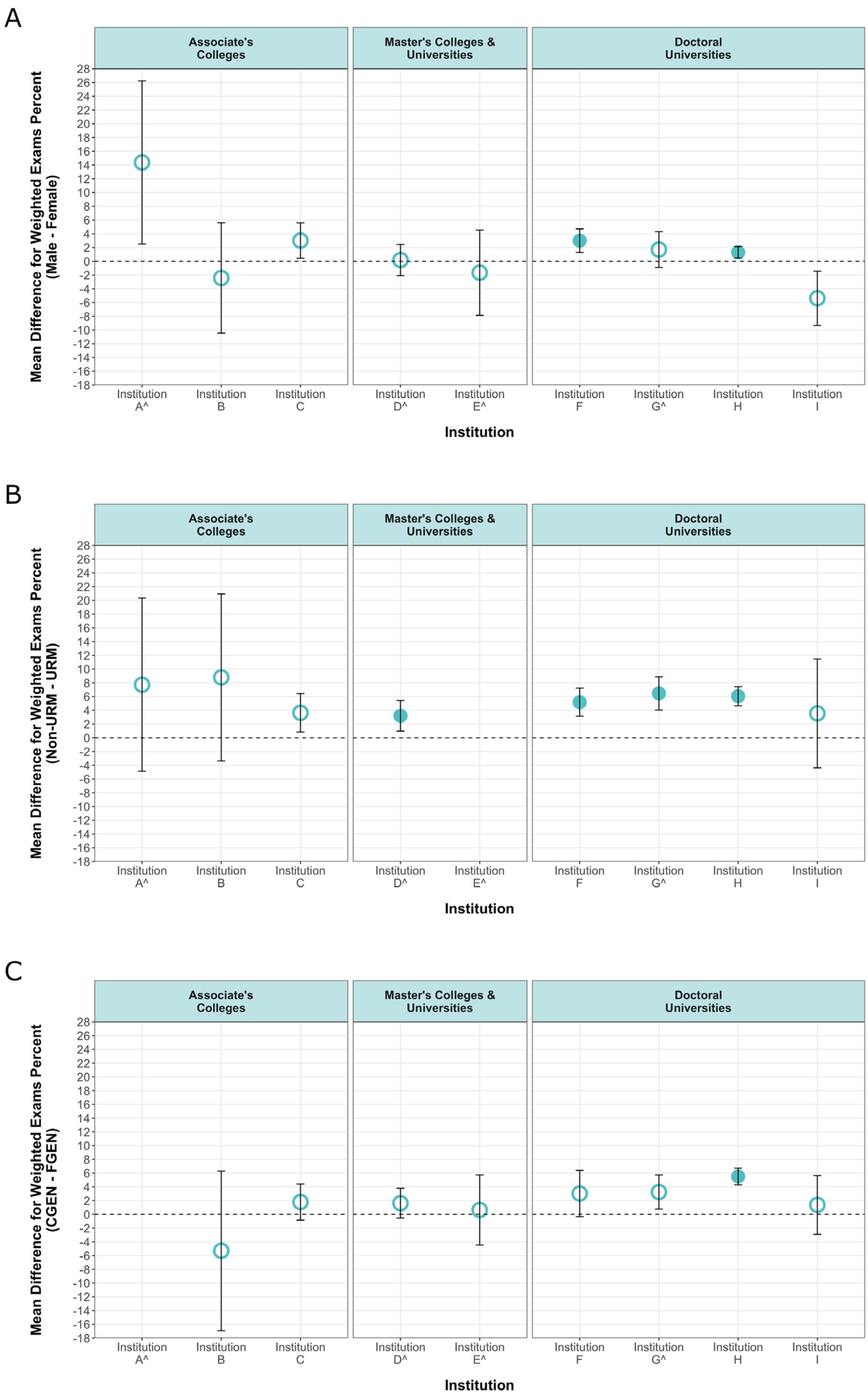

FIGURE 3. **Mean differences for weighted exam performance across institution type comparing (A) males and females, (B) non-URM and URM students, and (C) CGEN and FGEN students. In each panel, circles represent the differences between group means, and bars represent 95% confidence intervals for the differences between group means. Open circles indicate no significant difference between group means, and filled circles indicate a significant difference between group means. The dotted line represents no difference between groups. Measures above the dotted line indicate (A) males outperform females; (B) non-URM students outperform URM students; and (C) CGEN students outperform FGEN students. MSIs are designated with a ^.**

negative stereotypes about their academic ability, social isolation, and sexism from other students and faculty (Hill *et al.*, 2010; Robnett, 2016). They also report lower STEM self-efficacy than their male counterparts, even when actual academic performance is held constant (Robnett and Thoman, 2017; Marshman *et al.*, 2018).

We also realize the importance of institutional factors beyond the broad categorizations we mention here. Institutional transformation to evidence-based teaching and learning that is inclusive and equitable requires a systems-level analysis of the current behavior of the institution to understand: 1) where the problems lie, and 2) what changes are necessary to realize inclusive transformation. As faculty and institutions strive to improve biology education, we will use Nadler and Tushman's (1980) congruence model of organizational behavior as a framework to guide and evaluate institutional change as it occurs on the campuses of this network. The model posits that high congruence, or fit, among four factors that make up an organization—the *task*, the *people*, the *formal organizational structure*, and the *culture*—will positively impact behavior and performance. Thus, when observed outcomes do not align with desired outcomes, analyzing the congruency between the four components provides insight into areas within the organization where changes need to be made. By using the congruence model to understand where problems lie, RCN institutions can implement changes that are necessary to realize desired outcomes.

Intersectionality in Higher Education. Women of color encounter a combination of the aforementioned challenges as well as unique challenges that cannot be understood through their ethnicity or gender alone (Ong *et al.*, 2011; Williams *et al.*, 2014). The concept of intersectionality provides a framework for understanding these challenges. Some scholars argue that attaining a deep understanding of inequities in STEM fields requires consideration of the ways in which social categories combine to create distinct identity configurations (e.g., Ong *et al.*, 2011). In an undergraduate biology course, for example, a Latinx woman and a white woman may both encounter challenges related to gender, but the specific nature of these challenges and their implications may differ in meaningful regards.

EDU-STEM is well positioned to build on past work that applies the concept of intersectionality to the study of STEM disparities. Beyond diversity in the students and institutions comprising the sample, diversity within EDU-STEM itself (i.e., in terms of sociodemographic background and home institution) allows for multifaceted input into which research questions to prioritize and how to interpret core findings. We elaborate on key research priorities related to intersectionality that surfaced during our meeting.

Most of the existing research that applies an intersectional framework is qualitative and relies on small sample sizes from single institutions. Although this work is important and useful in its own right (e.g., Carlone and Johnson, 2007), it is also critical for larger quantitative studies to be included in the intersectionality literature. For instance, relative to qualitative research, quantitative studies allow for more formal hypothesis testing, the ability to statistically control for potential confounds, and clearer insight into the magnitude of group differences. Prior research is also limited, in that it has not made use of the full power of intersectionality. Specifically, a fair number of qualitative studies have focused on women of color in STEM (for a review, see Ong *et al.*, 2011), but little is known about how their experiences compare with the experiences of students from other backgrounds (e.g., men of color, white women). When such comparisons are conducted, sample size requirements often limit researchers to coarse ethnic groupings (e.g., lumping all URM students into the same category) and make it difficult to take into account more than two dimensions of identity (e.g., Robnett *et al.*, 2019).

The EDU-STEM data can be used in several ways to address these limitations. One possibility is a quantitative "deep dive" into the experiences of students who have often been overlooked in smaller-scale studies. For example, it would be worthwhile to examine whether the factors that predict academic success among Latinx men differ depending on whether the men are FGEN versus CGEN college students. Alternatively, the EDU-STEM data could also be used to cast a wider intersectional net. For example, we could compare mean levels of self-efficacy across all possible configurations of ethnicity, gender, and class background. This would provide insight into whether commonplace research findings (e.g., the finding that women have lower self-efficacy than men) hold across more complex identity configurations. More broadly, the EDU-STEM data can provide insight into how various facets of identity interact with the institutional context to shape student outcomes. For example, an African-American woman may have qualitatively different experiences in her biology class depending on whether she is enrolled at a PWI versus a historically black college or university.

It is important to emphasize, however, that EDU-STEM needs to be wary of reducing complex identity configurations to statistical interaction terms. We need to be mindful of the ways in which identity interacts with currents of privilege and power in the broader social context. Relatedly, at its core, the intersectionality framework is oriented toward fostering change by equalizing power imbalances that are often obscured by less-nuanced approaches. In this regard, the concept of intersectionality has clear implications for academic interventions that aim to reduce sociodemographic disparities in STEM engagement and performance. For example, if a woman of color and a white woman experience distinct challenges in their biology course, it follows that they may benefit from different types of interventions. Thus, in addition to documenting how student experiences vary at the intersection of multiple social categories, the intersectionality framework can help educators and researchers move beyond one size fits all interventions by informing the development of targeted interventions that optimize success for all students.

How to Grow the Network: A Recruitment Plan to Further the Goals of EDU-STEM

EDU-STEM was founded on the principle of increasing the literature representation of the student experience. Therefore, a major outcome of our meeting was the development of a recruitment plan for growing the network in the future. EDU-STEM aims to grow its membership by leveraging the personal relationships of network members to integrate new participants within existing institutional partnerships while broadening the network to include new institutional partners. Using a reciprocal partnership model as our guiding framework, EDU-STEM aims to cultivate a community that honors the contribution of all members and acknowledges the strengths and expertise that each partner brings to the table. In this sense, EDU-STEM is not about providing RIs with access to student populations at other institution types, instead it is about building relationships that allow different partners to best capitalize on the shared expertise and resources of the EDU-STEM membership. By involving a range of educators and researchers from a diverse set of institutions in collecting their own data on equity in STEM, this network will establish tendrils of equity awareness among groups of faculty not typically engaged in educational research or evidence-based teaching techniques. To achieve this, we will focus on two recruitment aims: 1) *developing institutional capacity* by recruiting a team of network participants from each partner institution and 2) *growing institutional representation* by recruiting new network participants from institutions that have been underrepresented in the literature.

Developing Institutional Capacity. Taking on the work of educational reform can be challenging if it is seen as a solo effort. EDU-STEM hopes to alleviate some of that challenge by cultivating communities of committed instructors within our partner institutions. By identifying an institutional lead (or two) at each of our partner institutions and then providing support for those leads to engage with other members of their institutions, EDU-STEM aims to support local communities that collectively contribute to the national group. EDU-STEM supports this local cultivation by providing travel support to meetings for network partners, curating opportunities for professional development around issues of equity and inclusion, and designing classroom-level interventions that are supported by the national network and available for all network partners to implement in their classrooms.

TABLE 3. Specific ways in which collaborative networks can promote and maintain partnerships with CCs and MSIs, institutions that are historically underrepresented in BER

Core network principles	Recommended practices
Authentic partnerships Partnerships should be based on opportunities for collaborations that are mutually beneficial for all participants.	Enable reciprocal exchange of ideas that create new value together rather than a transfer of resources from one partner to another. Deconstruct hierarchies to create opportunities for meaningful participation from multiple contexts. Facilitate opportunities for structured dialogue and shared learning to promote a commitment to creating common understandings. Provide pathways for constructive feedback and establish shared norms for giving and receiving feedback.
Targeted recruitment of underrepresented communities Facilitate targeted recruitment efforts for underrepresented communities by building relationships built on trust and shared commitments.	Institutional leaders should recruit from within their own institution to create a local community of support that can contribute to the broader network. Put personal relationship building at the front of conversations on partnership and emphasize shared ownership to promote trust. Personalize recruitment efforts to highlight the specific value added by a potential network member, including a commitment to shared values.
Public campaign to broaden access to network activities Make the network activities widely known by presenting at meetings, establishing a consistent brand identity, and maintaining a Web presence.	Make it easy for others to find out about you and your work. Establish a consistent brand identity and provide network members with recruitment materials (business cards, flyers, slides, etc.) that can be easily distributed to broad audiences. Host or sponsor professional development opportunities that build capacity and broaden knowledge for network participation. Have an application process in place to ensure that network membership remains in line with the network principles.
Incentivize and support network participation Honor the time and effort committed by network members in ways that meet their professional and personal needs.	Create clear guidelines for opportunities to participate in collaborative manuscripts, grants, meetings, and workshops. Provide funding support for participating in network events and create a process for the equitable distribution of available funds. Host network activities in a variety of locations to encourage participation from a greater number of network members and promote shared ownership. Use the network to lift up and advance the work of members in career stage–relevant ways; promote leadership opportunities for early-career researchers and students; and cultivate professional networking.

Inclusive Network Growth. How do we grow a research network with institutions that have been historically underrepresented in the literature? *Diversity* in science refers to cultivating talent and promoting the full inclusion of excellence across the social spectrum, including people from backgrounds that are traditionally underrepresented (Gibbs, 2014). From the perspective of inquiry, given that CCs and MSIs are leaders in successfully graduating URMs in STEM as well as matriculating them into graduate STEM academic programs (Fiegener and Proudfoot, 2013), few studies have been published that show or explain what these institutions are doing to be successful at educating underrepresented students. Via network involvement, CC and MSI faculty can contribute to emerging literature on promoting equitable participation in STEM. In the meeting, we discussed ways that the network can foster authentic, equitable partnerships with leaders at CCs and MSIs (outlined in Table 3) to better support the work of these scholars and provide opportunities for leadership that are responsive to the unique needs of these institutions.

Authentic Partnership. A core belief of the EDU-STEM network is that an effective network offers the opportunity for the creation of authentic partnerships, involving collaborations that mutually benefit the participants and participants' institutions and that create value together. Equitable participation can be enhanced via transparency at each step of the process. To that end, EDU-STEM participants decided to create and submit to consensus the following items in support of the network:

- a Principles of Operation document, clarifying shared terminology (e.g., STEM, MSI, PWI, gender, etc.)
- an authorship rubric, establishing criteria for involvement and allowing individuals to commit to different roles during manuscript development
- a project-submission process, whereby network members can "plant a flag" in a particular, specific, line of inquiry that draws on EDU-STEM data; any network member may join any project and contribute to resulting manuscripts, a formal process for proposing projects will prevent unnecessary duplication of efforts, provide members with a known point of contact, and promote accountability
- working groups, open to all members, with a specific charge, rotating leadership, and annual goals and objectives (e.g., data-management working group, network expansion working group)

By democratizing the organizational structure and encouraging participation that best leverages the experience, skills, and commitments of each individual member, EDU-STEM hopes to grow a collaborative network in a way that can promote equitable collaborations and sustainable partnerships.

Targeted Recruitment Activities. The goal of targeted recruitment within EDU-STEM is to ensure that traditionally excluded communities have access to network activities. During our meeting, there was a lengthy discussion about the need to recruit members in a way that is not exploitative and emphasizes trust, given that many minoritized populations have been subjected to exploitative relationships in the past. To this end, we have codified relationship building as the primary method for growing partnerships and encouraged collective decision making. We have asked our institutional leaders to serve as advocates within their own communities and to work to communicate with institutional leaders by sending emails, letters, brochures, and posters to deans and department chairs at MSIs, CCs, and other underrepresented institutions to recruit participants. By empowering and supporting current network members to take on leadership for targeted recruitment efforts, we hope to enhance the authenticity and credibility of efforts working toward a more inclusive STEM community.

Publicize the Network. In addition to our targeted recruitment, establishing an application process for a broader-based recruitment effort was a key priority coming out of the inaugural meeting. To achieve this, we proposed a three-pronged approach. First, we plan to recruit at meetings and poster sessions at local, regional, and national conferences. We will ask current network participants to represent the network at meetings they attend, particularly those attending meetings with a high attendance of underrepresented institutions and students (i.e., Annual Biomedical Research Conference for Minority Students, Southern Region Education Board, Emerging Researchers National Conference in STEM, Society for Advancement of Chicanos and Native Americans in Science).

In addition to recruitment efforts at meetings, we plan to host professional development opportunities that help build capacity and broaden the knowledge base of BER and broaden participation in research for BER faculty and non-BER faculty at MSIs and CCs. These efforts will include EDU-STEM participants giving guest lectures and EDU-STEM–designed professional development opportunities (workshops/seminars, virtual or in person). Through these supported opportunities, EDU-STEM will help cultivate a community of faculty from a variety of instruction types that have both the training and support to increase the implementation of evidence-based pedagogy and interventions.

Finally, EDU-STEM has partnered with the University of Minnesota's Impact Exchange (http://z.umn.edu/impactexchange) to develop a consistent network brand identity and online Web presence. Through this partnership, we will work with an undergraduate student intern who is receiving professional development in science communication and design through the Impact Exchange to create professional-quality recruitment materials (i.e., EDU-STEM give-aways) and to centrally manage the network's social media and website to inform potential participants about our network and maintain consistent network branding and messaging. We believe this will be an important part of developing a sustainable recruitment strategy, because it will allow for all network members to speak about the network in a consistent way and will empower members to take on recruitment efforts knowing they have the support of curated materials provided to them.

Incentivize Network Participation. Perhaps most important to our recruitment efforts and support of EDU-STEM is the acknowledgment that a traditional approach of relying on completely volunteer participation in activities will systematically exclude those with more limited access to resources. To move away from this model and to promote as much access to our network activities as possible, we are committed to providing access to resources that facilitate the dissemination of participant and network products, such as travel funds to participate in educational conferences and funds for the dissemination of educational publications. We are committed to leveraging network resources to particularly support members at critical career transitions (e.g., senior graduate students, postdocs, pre-tenure faculty). With access to network expertise and collaborations, members can leverage their network participation to successfully navigate key career transitions. Through our coordinated data-collection and intervention efforts, we anticipate that there will be huge potential for research output from the network, both through collaborative research proposals and collaborative manuscripts. Our goal is to make these activities accessible to all members of the network, particularly for those who may need to rely on such collaborations to be active/successful participants in BER.

Call for New Participants. If you have a passion for educational reform to promote more equitable STEM disciplines and want to get involved with the EDU-STEM network, we want to hear from you! You can find more information about the network and fill out an interest form on our website (edustemresearch.com). There you will find information about the current network members, ongoing network activities, and a link to sign up for the EDU-STEM newsletter. While the current focus of the network is confined to biology curricula based on the expertise of current network members, the future growth of the network welcomes participants from any STEM discipline. If you are looking to get involved or have any questions, reach out to us at edustemcontact@gmail.com, follow us on Twitter @EDUSTEMNetwork, or apply for network membership using this Google form (https://docs.google.com/forms/d/e/1FAIpQLSeV6mBgd1cTCB6ya85buWe9TFjWNm3DGXwSsLOeE_dW4C_I8A/viewform). We have also included an example of a recruitment letter in the Supplemental Material. Individual partners can best be reached through their institutions.

EDU-STEM members will be present at several national and international meetings over the coming years, so be sure to ask members about the network. Also, we will host an annual meeting each year directly preceding the Society for the Advancement of Biology Education Research meeting in July. If you are interested in attending the annual meeting, contact us via the webform and we will make sure you get added to the mailing list. Finally, we are putting together regular opportunities (semiannual to quarterly) for network members to connect virtually to continue conversations on projects and interventions.

CONCLUSION

It is important to note that EDU-STEM is not the only collaborative group working to promote equity and inclusion in undergraduate STEM. Additional examples include (but are not limited to) the Accelerating Systemic Change Network, the Association of American Universities Undergraduate STEM Education Initiative, the iEmber Network, the Association of American Colleges and Universities Undergraduate STEM Education initiative, and the Howard Hughes Medical Institute

Inclusive Excellence community. Through inclusive collaboration, we will gain insight into perspectives and address questions that would not be achievable otherwise. The Equity and Diversity in Undergraduate STEM meeting explored ideas about the present state of equity in undergraduate biology education, the largest barriers faced by institutions, and how a large-scale collaborative can contribute to the development of solutions through data generation and experimental efforts. We welcome interest from all members of the community and look forward to hearing from you!

ACKNOWLEDGMENTS

Samantha Brandt at University of Minnesota provided invaluable logistical and creative support for the meeting. We would also like to thank the Psychology Department at University of Nevada, Las Vegas, for hosting our meeting. This work was supported by a RCN grant from the NSF (DBI-1919462). Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

Aguillon, S. M., Siegmund, G.-F., Petipas, R. H., Drake, A. G., Cotner, S., & Ballen, C. J. (2020). Gender differences in student participation in an active-learning classroom. *CBE—Life Sciences Education*, *19*(2), ar12.

Allen, W. (1992). The color of success: African-American college student outcomes at predominantly white and historically black public colleges and universities. *Harvard Educational Review*. https://doi.org/10.17763/haer.62.1.wv5627665007v701

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC.

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

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.

Beckerson, W. C., Anderson, J. O., Perpich, J. D., & Yoder-Himes, D. (2020). An introvert's perspective: Analyzing the impact of active learning on multiple levels of class social personalities in an upper level biology course. *Journal of College Science Teaching*, *49*(3), 47–57.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, *44*(8), 1187–1218.

Carnegie Classification of Institutions of Higher Education. (n.d.). *About Carnegie Classification*. Retrieved April 23, 2020, from http://carnegieclassifications.iu.edu/

Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields* (Statistical Analysis Report NCES 2014-001). Washington, DC: National Center for Education Statistics.

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.

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable? *PLoS ONE*, *12*(12). https://doi.org/10.1371/journal.pone.0189610

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, *1989*, ar8.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2002). A temporal investigation of factors related to timely degree completion. *Journal of Higher Education*, *73*(5), 555–581. https://doi.org/10.1353/jhe.2002.0042

Diekman, A. B., & Steinberg, M. (2013). Navigating social roles in pursuit of important goals: A communal goal congruity account of STEM pursuits. *Social and Personality Psychology Compass*, *7*(7), 487–501.

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education*, *13*(3), 453–468.

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.

Espinosa, L. L., Turk, J. M., & Taylor, M. (2017). Pulling back the curtain: Enrollment and outcomes at minority serving institutions. Washington, DC: American Council on Education.

Fiegener, M. K., & Proudfoot, S. L. (2013). *Baccalaureate origins of US-trained S&E doctorate recipients*. Arlington, VA: National Science Foundation. Retrieved August 27, 2020, from https://nsf.gov/statistics/infbrief/nsf13323/nsf13323.pdf

Gasman, M., Baez, B., & Turner, C. S. V. (2008). *Understanding minority-serving institutions*. Albany, NY: SUNY Press.

Gibbs, K. (2014). Diversity in STEM: What it is and why it matters. *Scientific American*, *10*. Retrieved August 27, 2020, from https://blogs.scientificamerican.com/voices/diversity-in-stem-what-it-is-and-why-it-matters/

Gross, L. (2006). Scientific illiteracy and the partisan takeover of biology. *PLoS Biology*, *4*(5), e167.

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. https://doi.org/10.1126/science.1204820

Henning, J. A., Ballen, C. J., Molina, S., & Cotner, S. (2019). Hidden identities shape student perceptions of active learning environments. *Frontiers in Education*, *4*, 129.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, *466*(7302), 29.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Hernandez, P. R., Schultz, P. W., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology*, *105*(1). https://doi.org/10.1037/a0029691

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: AAUW.

hooks, b. (2000). *Feminist theory: From margin to center*. London: Pluto Press.

Johnson-Ahorlu, R. N. (2013). "Our biggest challenge is stereotypes": Understanding stereotype threat and the academic experiences of African American undergraduates. *Journal of Negro Education*, *82*(4), 382–392.

Lee, J. M., Jr, & Ransom, T. (2011). The educational experience of young men of color: A review of research, pathways and progress. *College Board Advocacy & Policy Center*. Retrieved August 27, 2020, from https://eric.ed.gov/?id=ED521425

Lohfink, M. M., & Paulsen, M. B. (2005). Comparing the determinants of persistence for first-generation and continuing-generation students. *Journal of College Student Development*, *46*(4), 409–428.

Ma, J., & Baum, S. (2016). Trends in community colleges: Enrollment, prices, student debt, and completion. *College Board Research Brief*, *4*, 1–23.

Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm? *Physical review physics education research*, *14*(2), 020123.

Massey, D. S., & Fischer, M. J. (2005). Stereotype threat and academic performance: New findings from a racially diverse sample of college freshmen. *Du Bois Review*, *2*(1), 45–67. https://doi.org/10.1017/S1742058X05050058

Mervis, J. (2011). Weed-out courses hamper diversity. *Science*, *334*(6061), 1333.

Nadler, D. A., & Tushman, M. L. (1980). A model for diagnosing organizational behavior. *Organizational Dynamics*, *9*(2), 35–51.

National Academies of Sciences, Engineering, and Medicine. (2019). *Minority serving institutions: America's underutilized resource for strengthening the STEM workforce*. Washington, DC: National Academies Press.

Ong, M., Wright, C., Espinosa, L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harvard Educational Review*, *81*(2), 172–209.

Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). Retrieved August 27, 2020, from https://eric.ed.gov/?id=ED338122

Rainey, K., Dancy, M., Mickelson, R., Stearns, E., & Moller, S. (2018). Race and gender differences in how sense of belonging influences decisions to major in STEM. *International Journal of STEM Education*, *5*(1), 10.

Robnett, R. D. (2016). Gender bias in STEM fields: Variation in prevalence and links to STEM self-concept. *Psychology of Women Quarterly*, *40*(1), 65–79.

Robnett, R. D., Nelson, P. A., Zurbriggen, E. L., Crosby, F. J., & Chemers, M. M. (2019). The form and function of STEM research mentoring: A mixed-methods analysis focusing on ethnically diverse undergraduates and their mentors. *Emerging Adulthood*, *7*(3), 180–193.

Robnett, R. D., & Thoman, S. E. (2017). STEM success expectancies and achievement among women in STEM majors. *Journal of Applied Developmental Psychology*, *52*, 91–100.

Sáenz, V. B., & Ponjuan, L. (2011). Men of color: Ensuring the academic success of Latino males in higher education. *Institute for Higher Education Policy*. Retrieved August 27, 2020, from https://eric.ed.gov/?id=ED527060

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. *Frontiers in Education*, *4*, 107. https://doi.org/10.3389/feduc.2019.00107

Schinske, J. N., Balke, V. L., Bangera, M. G., Bonney, K. M., Brownell, S. E., Carter, R. S., … & Fletcher, L. (2017). Broadening participation in biology education research: Engaging community college students and faculty. *CBE—Life Sciences Education*, *16*(2), mr1.

Setren, E., Greenberg, K., Moore, O., & Yankovich, M. (2019). Effects of the flipped classroom: Evidence from a randomized trial. *(EdWorkingPaper: 19-113)*. Retrieved from Annenberg Institute at Brown University http://www.edworkingpapers.com/ai19-113

Seymour, E., & Hewitt, N. M. (1994). *Talking about leaving: Factors contributing to high attrition rates among science, mathematics & engineering undergraduate majors: Final report to the Alfred P. Sloan Foundation on an ethnographic inquiry at seven institutions*. Boulder, CO: Ethnography and Assessment Research, Bureau of Sociological Research.

Snyder, J. J., Sloane, J. D., Dunk, R. D. P., & Wiles, J. R. (2016). Peer-led team learning helps minority students succeed. *PLoS Biol*, *14*(3), e1002398.

Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychology*, *52*(6), 613–629. Retrieved August 27, 2020, from www.ncbi.nlm.nih.gov/pubmed/9174398

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

Trapani, J., & Hale, K. (2019). *Higher education in science and engineering: Science and engineering indicators 2020* (NSB-2019-7). Arlington, VA: National Science Foundation.

Williams, J., Phillips, K. W., & Hall, E. V. (2014). *Double jeopardy? Gender bias against women of color in science*. San Francisco: Hastings College of the Law, Center for WorkLife Law.

Wilson, Z. S., Holmes, L., Sylvain, M. R., Batiste, L., Johnson, M., McGuire, S. Y., … & Warner, I. M. (2012). Hierarchical mentoring: A transformative strategy for improving diversity and retention in undergraduate STEM disciplines. *Journal of Science Education and Technology*, *21*(1), 148–156.

Appendix 3. Half-century of student data reveal professional impacts of a biology field course (In review, formatted for *Proceedings of the National Academy of Sciences*)

Contribution: Collected data; Performed qualitative analysis; Revised and edited drafts of the paper.

Half-century of student data reveal professional impacts of a biology field course

Lina M. Arcila Hernández[1,*], Cinnamon S. Mittan[2,*†], Todd Lamb[3,¶], Katherine D. Holmes[4,*], Coby McDonald[5,*‡], Kelly R. Zamudio[6,*§], Cissy J. Ballen[7,¶]

Affiliations:
*. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14850
† .Kellogg Biological Station, Michigan State University, Hickory Corners, MI
‡. Department of Biology, Colorado State University, Fort Collins, CO 80523
§ .Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712
¶. Department of Biological Sciences, Auburn University, Auburn, AL 36849

Contact information:

1. lma84@cornell.edu
2. mittanci@msu.edu
3. tcl0011@auburn.edu
4. kdh73@cornell.edu
5. coby.mcdonald@colostate.edu
6. kelly.zamudio@austin.utexas.edu
7. mjb0100@auburn.edu

Corresponding authors: Lina M. Arcila Hernández and Cinnamon S. Mittan. Equal contributors

Keywords: ecology and evolutionary biology; graduate program; longitudinal dataset; biology education research; STEM careers

Manuscript for consideration as a Research Report in *Proceedings of the National Academy of Sciences*

Instructions: https://www.pnas.org/authors/submitting-your-manuscript#initial-submissions

Abstract

Field courses provide intensive learning experiences that diversify curricula and inspire students through immersion in non-traditional academic environments. Despite the potential benefits of field courses, there is little detailed information on the impact of field courses on student outcomes, the magnitude of these impacts, and how student participation may influence career trajectories. We used a nearly 50-year longitudinal dataset of graduate students in the same ecology and evolutionary biology program to compare career outcomes for students who participated in a field course versus those who did not. More broadly, we surveyed all students since 1960 to identify the graduate experiences and skills most useful to advancing students' scientific careers. We found that field course attendees co-authored more scientific publications compared to non-attendees, both during graduate school and up to ten years after graduation. While students in both groups graduated and continued on to scientific careers at similar rates, students attending the field course became faculty at a higher rate. Our survey data showed that field courses provide opportunities to engage with activities identified as critical to success in graduate school, including student-led environments (*i.e.*, peer feedback and teaching), and mentor feedback. Students also reported gaining important skills from field courses, such as carrying out research and learning from observing nature. Our work demonstrates that field courses impact student experiences and potentially alter career trajectories, underscoring their importance as effective pedagogical tools to train the next generation of scientists.

Abstract word count: 238

Significance Statement

Immersive field courses provide students with opportunities to think creatively and address research questions using data from the natural world. To understand the impact of field courses on biology graduate students, we harnessed a nearly 50-year longitudinal dataset to compare professional outcomes for students who attended a two-week field course to outcomes for students who did not attend. We identified a relationship between field course attendance and scientific career outcomes and publication outputs. Based on survey data, field course participants perceived carrying out research and learning from their observations of nature to be among the most important takeaways from their experiences. Our results highlight the importance of field courses in our national efforts to promote student learning and development of scientific innovation.

Introduction

Improving recruitment and retention of students in the science, technology, engineering and mathematics (STEM) workforce has critical consequences, as science impacts every corner of society and our daily lives. Field courses provide unique opportunities to engage and retain STEM students, as learning in field settings invokes a sense of wonder about the natural world and encourages a lifetime pursuit of science (Dayton and Sala, 2001; National Research Council 2014). Immersive field courses provide a powerful outlet for students to think creatively and ask questions using observational and experimental data gathered in an ecological context

(Fleischner et al., 2017; National Research Council 2014). As the ability to address research questions primarily with molecular data or mathematical modelling increases, providing opportunities for students to engage with research beyond laboratory or computational settings will be a challenge for graduate education. Immersive field courses offer an excellent opportunity for students to make observations of nature, collaborate with their peers, and experiment in field research.

Previous research reveals that field courses teach practical skills (Dillon et al., 2006) and increase retention and success among science majors, especially for students within historically underserved groups in science (Beltran et al., 2020; Mason et al., 2017). Despite documented benefits, university support for field courses and field stations is diminishing, such that future generations of biologists will experience research in biological sciences that is isolated from its ecological context (Smith 2004; Fleischner et al., 2017; National Research Council 2014).

More broadly, quantitative assessments of research experiences are limited (Fleischner et al., 2017; Linn et al. 2015), particularly in the context of field courses. This is due to several challenges, including the difficulty associated with tracking long-term scientific outcomes and small sample sizes. An additional difficulty in understanding the impacts of field courses is the observational nature of such studies. For example, validated student surveys that focus on field courses are uncommon in the education research literature. When present, surveys are not typically designed to evaluate generalizable impacts of field courses on student outcomes [Shinbrot et al., in review]. Furthermore, the majority of education research on field courses seems to be targeted to undergraduate, rather than graduate-level experiences (Leon-Beck & Dodick, 2012). To our knowledge, there has been no quantitative analysis of how exposure to field course experiences during graduate education impacts long-term student outcomes as scientists. Given national efforts to promote the retention of students in STEM, this fundamental topic highlights the importance of research that explores field course impacts on STEM students at all levels (Mason et al., 2018; Moore 2001).

We investigated the impacts of a field biology course on various quantitative and qualitative outcomes for Ecology and Evolutionary Biology graduate students who attended an annual Florida Field Course (FFC) offered by Cornell University, a large research institution in the northeastern United States. This course has been offered to graduate students near-continuously since 1972 with little change to course structure or curriculum, allowing comparison over the course's history. We use quantitative data collected over nearly five decades to measure progress in the graduate program (graduation rate), research productivity (authorship rates during graduate school and for 10 years after graduation), and career trajectory. We combine this with a qualitative survey of 131 former students to identify the activities and experiences most useful to their scientific training more broadly, and, among field-course attendees, the skills students acquired from their field course experience. Because field biology courses have been associated with increased retention in science majors (Beltran et al., 2020; Mason et al., 2017), we expected to see increased graduation rate and an increase in science careers for alumni of the FFC, as well as higher scientific publication rate among these students given the course requirement to complete an independent research project. To improve our collective approach to field education, we offer recommendations about the most important field course attributes so that future generations of biologists may benefit from them.

Results

Quantitative analyses

Our dataset consisted of all graduate students admitted into the Ecology and Evolutionary Biology program at Cornell between 1960-2016 (n = 459) (Table 1, Supplementary Information 1), as the FFC was offered through this program, and largely attended by its graduate students. We also included students from different departments at the same university (n = 144; total individuals from both sets n = 603) who took part in the FFC during the same period. We compared available data on student retention ("Retention dataset"), publication rates ("Publications dataset"), and career type ("Career dataset") from individuals who attended the field course and those who did not attend (Table 1). Retention rates in graduate school were generally high (93% graduation rate; 40 students did not graduate), and we were unable to detect a difference in retention rates between FFC and non-FFC students (Table 1, Chi-square = 1.48; p-value = 0.29). While participants in the field course were self-selected, we compare participant outcomes to similar students who did not participate in the field course (Table 1). Given the similarity in retention rate between the two groups, we assume that FFC attendees and non-attendees were equally motivated to complete their graduate education.

Table 1: General characteristics of the career trajectory, publications, and survey data sets for those that did, and did not attend the Florida Field Course.

	Attended FFC	**Did not attend FFC**	**Total**
Retention dataset	184 (31%)	408 (69%)	592
Career dataset	175 (34%)	329 (65%)	507
Publications dataset	139 (30%)	331 (70%)	470
Survey Responses	55 (42%)	76 (58%)	131
EEB Program	141 (31%)	317 (69%)	458
PhD Program	161 (35%)	301 (65%)	462
Average time to degree	6.45	6.27	-

We also examined publication rates to assess differences between FFC and non-FFC students during graduate training, and after graduation. We found that while in graduate school, FFC students published more papers (27% increase; n = 139, emmean $\pm$ se: 3.54 $\pm$ 0.66; p-value < 0.001) than students who did not attend the field course (n = 331, emmean $\pm$ se: 2.78 $\pm$ 0.51. Figure 1A). Ten years after graduation this difference persisted; students in the field course published about 14% more (emmean ± se: 10.04 ± 2.88; p-value < 0.001) than their peers who did not participate in the course (emmean ± se: 8.76 ± 2.51; Figure 1B).

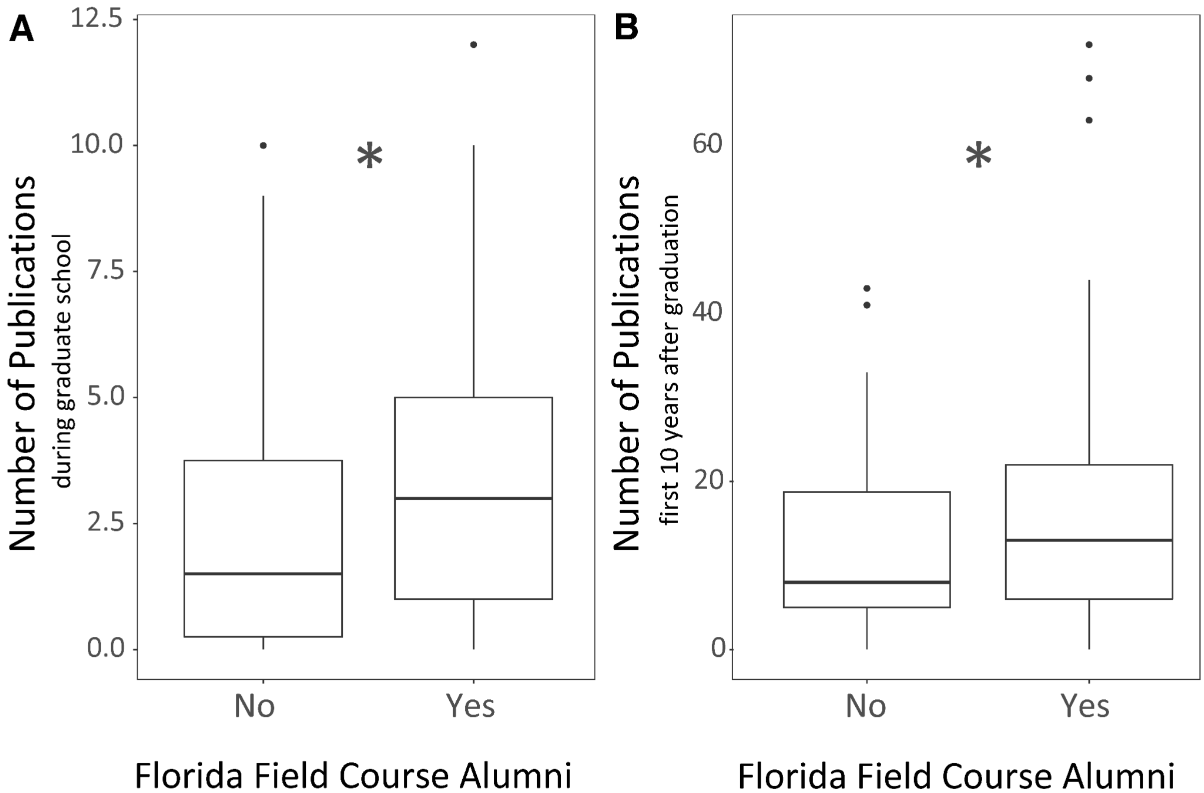

Figure 1. Number of scientific publications across graduate alumni who did not participate in the Florida Field Course (FFC) (No) and those who did participate (Yes); (A) Publications during graduate school; (B) Publications ten years after graduating.

Finally, we asked whether career types post-graduation differed between FFC attendees and non-attendees. We found that FFC students held a higher number of faculty positions in research institutions (p-value = 0.005) and potentially in teaching institutions (p-value = 0.07), though the difference is not significant at the 0.05 confidence level. However, the majority across both groups pursued careers in STEM, either as researchers in academic or non-academic settings (Figure 2).

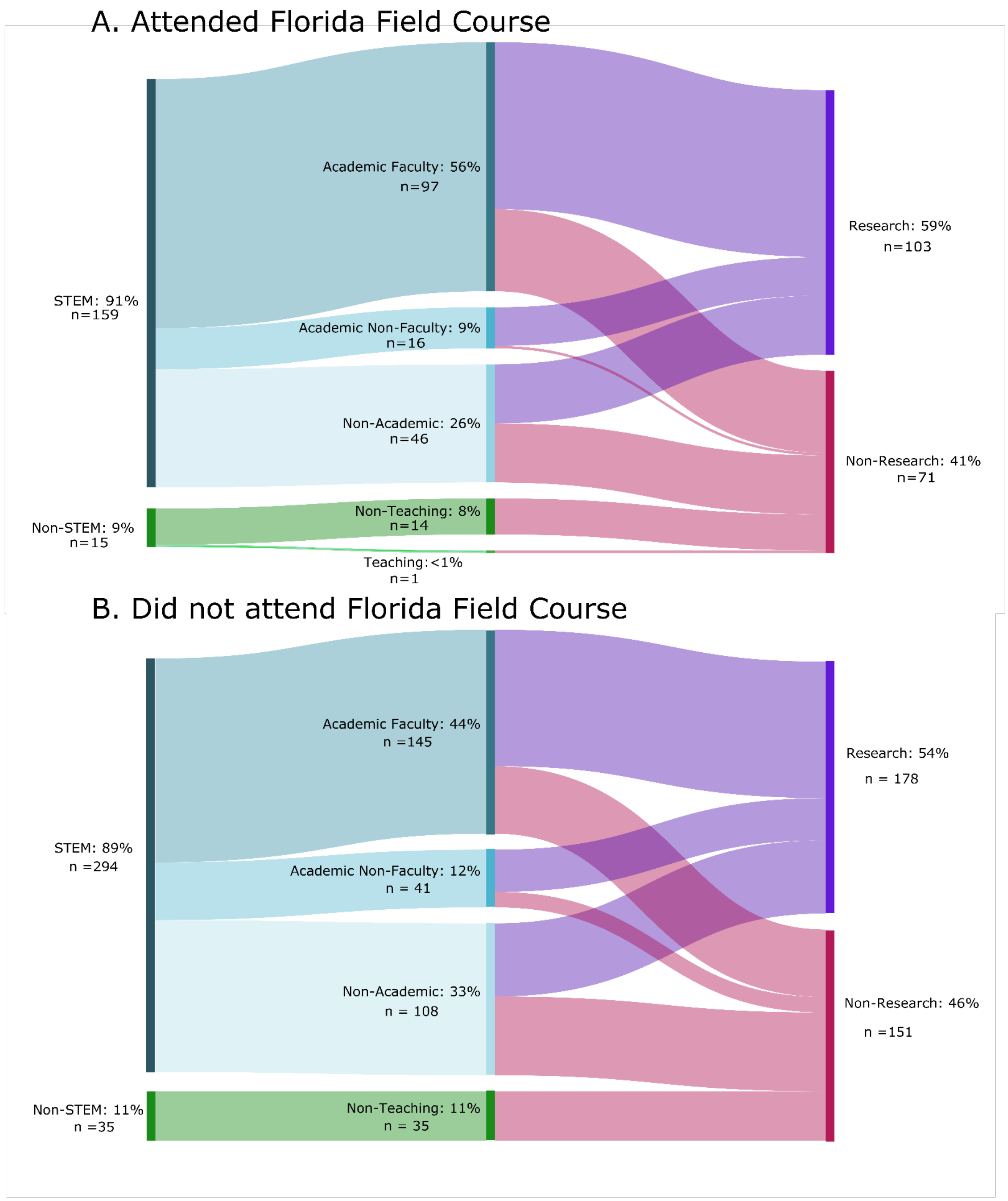

Figure 2. Career pursuits by category across graduate alumni who (A) attended the Florida Field Course (FFC) and (B) those who did not attend. Careers are first organised by STEM or Non-STEM on the left, further characterised into subcategories (i.e.- Academic or Non-academic) in the middle, and split by Research or Non-Research on the right.

Qualitative analyses

We used the software platform Qualtrics (www.qualtrics.com) to administer a survey (see Supplemental Information 3) to everyone enrolled in the graduate program since 1960 and for whom we had email addresses (n= 447) to determine which experiences or skills were most relevant to their scientific career. We received completed surveys from 131 alumni. We conducted qualitative analysis to interpret two open-ended response items from the survey. First, we asked all graduate students (*i.e.*, those who did and did not attend the FFC) which early experiences in graduate school were most useful for analysing data and writing scientific papers (Table 2). The most frequently identified experiences were student-led environments (33%), mentor feedback (32%), and the graduate classes (27%). Field courses were also reported as a useful experience for scientific writing (4%).

Table 2. The most commonly reported experiences in graduate school that students found useful for analysing data and writing scientific papers. Note that participant responses could include more than one theme.

Theme	**Percentage**
Student-Led Environments	33%
Mentor Feedback	32%
Graduate Classes	27%
Required Task/Independence	21%
Seminars	11%
Grant Application	10%
Collaboration	9%
Statistics Consulting Unit	4%
Field Course	4% *

*(only participants that attended a field course: 5%)

Second, we asked FFC attendees to describe the most impactful learning or skill they took away from their field course experience (Figure 3). The top three reported skills were: the opportunity to carry out research in the field (43%); thinking broadly about science, curiosity, or discovery, such as exposure to new disciplines and approaches to research (28%); and third, learning from and observing nature (25%).

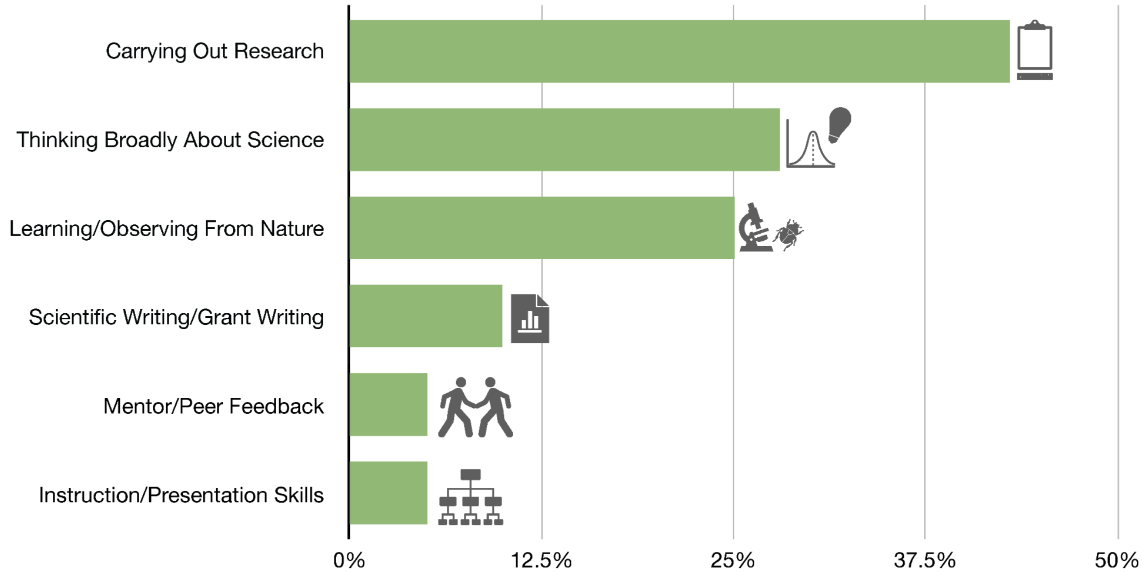

Figure 3. The most commonly reported learning experience or skill that graduate students took away from a field experience. Percentages reflect survey participants whose responses fit into a corresponding theme.

Discussion

The current study represents the most comprehensive work to date on the long-term impacts of biology field courses on graduate students. Using a 44-year longitudinal dataset, our results show that student participation in an immersive field course provides skills and experiences aligned with the current aims of the graduate program, such as an increase in publication rates and persistence in a research, environmental policy, or teaching career. Participant survey responses indicated student-led environments and mentor feedback were among the most relevant to career success (Table 2 and Figure 3). In addition, of those students who reported attending any field course in graduate school, 5% mentioned them as most useful for analysing data and writing papers. Further, we had not asked participants about field courses at this point in the survey, thus, participants self-reported the impact of field courses without prompting. This feedback is broadly applicable to the design of graduate programs, as well as how to strengthen field courses to best prepare students. Alumni who attended the field course expressed that its most important feature was the backdrop of nature, in which they could observe and experiment with research. This result aligns with the theory behind sense-of-place education, where an individual's relationship to their place of learning can be correlated with, and have a profound impact on, the individual's academic performance (Semken and Brandt, 2010; Johnson et al., 2020; National Research Council 2014). Our collective findings provide evidence of the overall benefits of field courses, suggesting that increasing both the number and accessibility of field courses has the potential to enhance innovation in biology.

The results of this study are compelling; however, we note that our data have several limitations. First, our analyses are correlative, and so we are unable to infer unidirectional effects of the field course on student outcomes. Second, students self-selected to attend the field course; thus, the participant pool may be biased towards individuals who are already positioned for success in academia or who are targeting faculty positions as an end goal. Nonetheless, we found that the FFC attendees and non-attendees were similar in many characteristics, including graduation rate, time-to-degree, and pursuit of STEM careers. Similarly, we were unable to assess academic preparation prior to graduate school as our survey only assessed graduate school experiences. To address these limitations, we used cohort as a random variable to explain some of the variation across years, such as previous preparation, which we expect to be more similar within cohorts. Third, our study conclusions are limited by the demographic representation within the field-course attendees, and graduate students more generally. While undergraduate participation in field courses can decrease achievement gaps in Ecology and Evolutionary Biology, graduate-level programs in Ecology and Evolutionary Biology are among the least diverse in STEM (Graves, 2019; O'brien et al., 2020; Massey et al., 2021). Thus, evaluating the potential impact of field courses on retention of Historically Underrepresented Minorities in graduate education is an important gap in education literature.

Our results show long-term impacts on student outcomes, in alignment with the few existing studies documenting the impacts of field courses on short-term student outcomes (Beltran et al., 2020; Cotton, 2009; Easton and Gilburn, 2012). Given these potential long and short-term impacts, it is important to address the barriers preventing students from attending field courses. Students may be reluctant or unable to spend time away because of family or work obligations. While field courses can be carried out at any time of year, many take place outside of semester teaching hours, or during holidays and weekends, making it difficult for individuals who have other obligations during those periods (Smith, 2004; Fleischner et al., 2017). Additionally, field courses may not be inclusive to students on the basis of ability, gender, race, LGBTQIA-status, socioeconomic status, or other identities currently and historically excluded from science. Advocating for accessibility to field courses for such students will foster a culture of inclusion in natural sciences (Carabajal et al., 2017; Gilly et al., 2015; Pickrell 2020).

We offer empirical evidence of long-term positive associations of a field-based education experience on student outcomes (*e.g.*, sense of curiosity, and rate of scientific publications) in a biology graduate program. In light of these results, as well as previously reported enthusiasm for field courses from educators (Beltran et al., 2020; Fleischner et al., 2017) and from students at multiple age levels (Barker, Slingsby and Tilling 2002; Boyle et al. 2003, 2007; Cotton and Cotton 2009; Goulder et al., 2013), we encourage continued investment in field programs. Moreover, the recent trend towards more equitable student-centered active learning pedagogy (Freeman et al., 2014; Theobald et al., 2020) and authentic research experiences in biology (Bangera & Brownell 2014) has intensified focus on the potential value of field-based education opportunities. Describing the impacts of field courses should strengthen understanding and support among students, colleagues, and administrators in higher education.

Methods

Field Course and Student Population

The Graduate Field Course in Ecology at Cornell University, best known as the Florida Field Course (FFC) is an ideal program to examine the impact of field courses on student outcomes in higher education. It provides a highly structured experience for graduate student cohorts over 15 days in the spring, during which students independently develop research projects and collaboratively work with peer scientists. Students also participate in a variety of professional development enrichment activities, such as learning to read research articles, pitching their research idea to peers, and presenting conclusions through oral presentations and written manuscripts. The goal of the course is to immerse students in an ecological system where they can discover opportunities for exploration and address biological questions.

The FFC began in 1968 when professor Dick Root travelled to Archbold Biological Station with a group of graduate students to research insect-plant interactions. The course has persisted for over 50 years, with the course structure remaining virtually unchanged. A central goal of the course is to support students in successfully developing and testing field-based research questions. In addition, students are encouraged to explore research systems and techniques beyond the sub-discipline of their graduate research, with the aim of broadening their scientific experience.The course is open to all graduate students but mostly advertised within the Cornell Department of Ecology and Evolutionary Biology and associated departments.

The course typically enrols approximately 20 students who travel to Archbold Biological Station for two weeks during Spring Break. At the station, students are taken to a variety of ecosystems over the first few days to become familiar with the local fauna and flora. They also interact with local researchers through guided nature walks and presentations that model field-based hypotheses and experimental design. Then, students develop research questions, test hypotheses with observational and experimental studies, share results in the form of oral presentations, and write first drafts of scientific articles. Several students published the manuscripts that were developed during the field course in peer-reviewed journals (e.g., Mason 2017 and Goud 2017).

Longitudinal Data set

To address our research questions, we collected the following information from the university registrar or public sources: email address, FFC alumnus (yes or no), year of FFC attendance, year graduate school enrollment, year of graduate school completion, degree type obtained (Masters or Doctoral), current institution (if applicable), and current career. We assigned each alumnus to one of nine career types based on a few themes of interest: research or non-research, academic or non-academic, STEM or non-STEM, faculty or non-faculty, and teaching or non-teaching (Table 4).

Table 4. The nine categories of career type used in analyses. Careers were classified by academic or non-academic, faculty or non-faculty, teaching or non-teaching, stem or non-stem, and research or non-research.

Category	Example
academic, faculty, teaching, stem, non-research	Faculty at a teaching-intensive institution
academic, faculty, teaching, non-stem, non-research	Faculty at a teaching-intensive institution in a non-STEM field
academic, faculty, teaching, stem, research	Faculty at a research-intensive institution

academic, non–faculty, non-teaching, stem, research	Research associate at a university
academic, non-faculty, teaching, stem, non-research	Lecturer in STEM field
academic, non-faculty, teaching, non-stem, non-research	Lecturer in non-STEM field
non-academic, stem, research	Government researcher
non-academic, stem, non-research	Director of a STEM-related NGO
non-academic, non-stem, non-research	Careers that do not fall under any of the previous themes

Through systematic searches, we also recorded the number of publications each individual co-authored during graduate school and the number of publications within the 10 years following graduation. We used the number of scientific publications as a proxy for scientific contributions during and immediately after graduation. To obtain these data, we searched two databases for each individual potential author from the program (See full protocol in Supplementary Information 2): Google Scholar and Web of Science. For each author, we refined the search by author name and date range. First, we searched for publications from when the individual started their graduate program to the year they graduated. Next, we searched from the year they finished the graduate program to 10 years after they finished the program. We searched in both databases and took the highest number of publications. For example, if Google Scholar found five publications during the graduate program, and Web of Science found three, we recorded five for the total number of publications during the student's graduate program.

Quantitative analyses

We compared performance outcomes of graduate school students who participated in the FFC (n = 184) to students in the same graduate program who did not participate in the field course (n = 408). We removed 53 entries from our original data set of 652 students because we did not have enough information about the year the students started or ended the program, or we were not able to find information on career type after the student ended the program.

We conducted all statistical tests in RStudio (Version 1.3.959). To determine graduation rates between the two groups of students, we ran a contingency test (*chisq.test*) using the libraries *psych (Version 1.9.12.31;* Revelle 2019*)*, *lsr (Version 0.5;* Navarro 2015*)*, and *car (Version 3.0-8;* Fox and Weisberg 2019*)* with 10000 replicates to compute simulated p-values in the Monte Carlo test. To determine if there were differences in the number of publications between students that took the field course experience and those that did not, we used a Generalized Linear Mixed Model using the library *glmmTMB* (Version 1.0.1; Brooks et al. 2017) (glmmTMB(formula = Publications ~ FFC Alumni + (1|cohort), data=dataframe, family = "poisson", na.action = na.exclude)). We assigned the number of publications during the degree program and 10 years after graduation ("Publications") as response variables and participation in the field course ("FCC Alumni": yes or no) as a fixed factor. Because the average time to graduation was 6.3 years, we were able to use this information (rounded to 6 years) to create 9 cohorts starting in 1960 (the year the first student in the data set enrolled in the graduate program) until 2016 (the last year a student enrolled in the graduate program in our dataset). On average, there were 60 students per cohort (see Supplementary Information 5 for counts per cohort).Thus, we used cohort as a random variable in our models. In this study, we report the Estimated Marginal Means (emmeans) and standard errors obtained with the *emmeans* package (Version 1.4.7; Lenth 2020).

Student Experience Survey
We designed a survey to determine the skills and experiences most relevant to the scientific careers of alumni (see questions, Supplementary Information 3). Prior to broadly distributing the final version of the survey, we sent a draft of the survey to five current and former biology graduate students and two qualitative researchers to obtain feedback on survey clarity, and to ensure that questions were interpreted as intended. We modified the survey language in response to their comments. We obtained publicly available email addresses of students in the graduate alumni and contacted individuals about participating in a survey about their graduate school experiences. We used Qualtrics to administer the survey to all former students enrolled in the graduate program since 1960 for whom we had email addresses (n= 447) to determine what experiences or skills were most relevant to their scientific careers.

We avoided priming participants about field course experience by broadly asking: "What year during your graduate school experience did you start independently asking scientific questions (i.e. first year, second year, etc)? Give specific examples (3-5 sentences)." By beginning the survey with this general question, we aimed to elicit holistic reflections on useful graduate school experiences that can be woven into field courses.We then asked participants: "write about your early experiences in graduate school analysing data and writing scientific papers. What experiences were most useful? Give specific examples (3-5 sentences)."

Following these open response questions, we asked participants to categorise which career best describes the pathway they took at the completion of their graduate training. After collecting this information, we asked participants if they attended the Florida Field Course, offered to all students in the graduate program. The survey concluded for those students who indicated they had not. Participants who indicated they had attended the FFC were directed to the last section of the survey. This section included open-ended questions, and we requested that participants write 3-5 short sentences reflecting on their field course experience. We asked "What was the most impactful learning or skill you took away from the program?" We also asked participants a few questions for internal program purposes, such as at what level they perceived was the best time to participate in the field program (e.g., first year, second year, etc.), what they would say to a graduate who was considering the program, what they would change or improve about the program, and whether there was anything else they would like us to know (see survey in Supplementary Information 3).

Qualitative analyses

We coded two open response questions from the student experience survey. For each of the open response questions, two researchers (TL and an undergraduate research assistant) coded the data to consensus, reducing the potential for variability among coders and minimising the likelihood of omitting critical categories. Coders created categories through first- and second-cycle analyses using emergent or inductive coding (Saldaña, 2021) (see Supplemental Information 4 for themes and coding rubric). We adapted themes to be increasingly inclusive and descriptive. We calculated percent agreement and percent exclusion on a per-question basis. We split responses into 40-50 "blocks" (*i.e.*, rows of responses). TL and an undergraduate research assistant coded these blocks separately and then compared their codes and achieved a Cohen's κ interrater score at an acceptable level (κ = 0.90; Landis and Koch, 1977). For instances in which the two coders could not reach consensus, an additional author (CJB) was consulted until consensus could be reached.

Acknowledgements
We are grateful to those who provided us with feedback on initial drafts of the survey, including Brie Craig and Sadie Hebert. We would also like to thank the biology education research group at Auburn University for valuable feedback on the manuscript at various stages: Emily Driessen, Abby Beatty, Sharday Ewell, Todd Lamb, Ash Zemenick; and undergraduate research assistants Rachel Youngblood and Priyanka Manon. This work was supported by the National Science Foundation (grants DBI-1919462, DUE-2011995, DUE- 2120934 awarded to CJB; Cornell Active Learning Initiative Fellowship to LMAH).

Land Acknowledgement
Cornell University is located on the traditional homelands of the Gayogo̱hó:nǫ' (the Cayuga Nation). The Gayogo̱hó:nǫ' are members of the Haudenosaunee Confederacy, an alliance of six sovereign Nations with a historic and contemporary presence on this land. The Confederacy precedes the establishment of Cornell University, New York state, and the United States of America. We acknowledge the painful history of Gayogo̱hó:nǫ' dispossession and honor the ongoing connection of Gayogo̱hó:nǫ' people, past and present, to these lands and waters. We also acknowledge the use of the traditional homelands of the Calusa and Seminole nations on which Archbold Biological Station is located.

Literature Cited
Barker, S., D. Slingsby, and T. Tilling. 2002. Teaching biology outside the classroom: Is it heading for extinction? Shrewsbury, UK: Field Studies Council

Bangera, G., and S.E. Brownell. 2014. Course-based undergraduate research experiences can make scientific research more inclusive. CBE—Life Sciences Education 13(4): 602-606.

Beltran, R. S., Marnocha, E., Race, A., Croll, D. A., Dayton, G. H., and E.S. Zavaleta. 2020. Field courses narrow demographic achievement gaps in ecology and evolutionary biology. Ecology and Evolution 10(12): 5184-5196.

Boyle, A., S. Conchie, S. Maguire, A. Martin, C. Milsom, R. Nash, S. Rawlinson, A. Turner, and S. Wurthmann. 2003. Fieldwork is good? The student experience of field courses. Planet Special Edition 5(11) 48–51.

Boyle, A., S. Maguire, A. Martin, C. Milsom, R. Nash, S. Rawlinson, A. Turner, S. Wurthman, and S. Conchie. 2007. Fieldwork is good: The student perception and the affective domain. Journal of Geography in Higher Education 31(2): 299–317.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A. Skaug, H.J., Maechler, M., and B. M. Bolker. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal 9(2): 378-400.

Carabajal, I. G., Marshall, A. M., and C.L. Atchison. 2017. A synthesis of instructional strategies in geoscience education literature that address barriers to inclusion for students with disabilities. Journal of Geoscience Education 65(4): 531-541.

Cotton, D.R.E., and A.A. Cotton. 2009. Field biology experiences of undergraduate students: The impact of novelty space. Journal of Biological Education 43(4): 169–74.

Dayton, P. K., and E. Sala. 2001. Natural history: the sense of wonder, creativity and progress in ecology. Scientia Marina 65(S2): 199-206.

Dillon, J., Rickinson, M., Teamey, K., Morris, M., Choi, M. Y., Sanders, D., and P. Benefield. 2006. The value of outdoor learning: evidence from research in the UK and elsewhere. School science review 87(320): 107.

Duc Bo Massey, M., Arif, S., Albury, C. and Cluney, V.A. 2021. Ecology and evolutionary biology must elevate BIPOC scholars. Ecology Letters, 24: 913-919. https://doi.org/10.1111/ele.13716

Easton, E., and A. Gilburn. 2012. The field course effect: gains in cognitive learning in undergraduate biology students following a field course. Journal of Biological Education 46(1): 29-35.

Fleischner, T. L., Espinoza, R. E., Gerrish, G. A., Greene, H. W., Kimmerer, R. W., Lacey, E. A., ... and L. Zander. 2017. Teaching biology in the field: Importance, challenges, and solutions. BioScience 67(6): 558-567.

Fox, J., and S. Weisberg. 2019. An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gilley, B., Atchison, C., Feig, A., and Stokes, A. 2015. Impact of inclusive field trips. Nature Geoscience 8(8): 579-580.

Goud, E.M. 2017. Diversity and abundance of litter-dwelling Arthropods increase with time-since-burn in a Florida scrub ecosystem. Biodiversity 18(4): 151-155.

Goulder, R., Scott, G. W., and L. J. Scott. 2013. Students' Perception of Biology Fieldwork: The example of students undertaking a preliminary year at a UK university. International Journal of Science Education 35(8): 1385-1406.

Graves, J.L. 2019. African Americans in evolutionary science: where we have been, and what's next. Evolution: Education and Outreach 12, 18. https://doi.org/10.1186/s12052-019-0110-5

Hernandez, P. R., Woodcock, A., Estrada, M., and P.W. Schultz. 2018. Undergraduate research experiences broaden diversity in the scientific workforce. BioScience 68(3): 204-211.

Landis, J. R., and G.G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics, 33(1): 159–174.

Leon-Beck, M. and J. Dodick. 2012. Exposing the Challenges and Coping Strategies of Field-Ecology Graduate Students, International Journal of Science Education, 34(16): 2455-2481.

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., and E. Stone. 2015. Undergraduate research experiences: Impacts and opportunities. Science 347: 6222.

Mason, N. A. 2017. Effects of wind, ambient temperature and sun position on damselfly flight activity and perch orientation. Animal Behaviour 124: 175-181.

Mason, N. A., Brunner, R. M., Ballen, C. J., and I.J. Lovette. 2018. Cognitive and Social Benefits among Underrepresented First-Year Biology Students in a Field Course: A Case Study of Experiential Learning in the Galápagos. Frontiers: The Interdisciplinary Journal of Study Abroad, 30(3): 1-19.

Moore, P. G. (Ed). 2001. Developing and sharing best practice in marine-related fieldwork. University Marine Biological Station, Millport.

National Research Council. 2014. Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the 21st Century. Washington, DC: The National Academies Press. https://doi.org/10.17226/18806.

Navarro, D. J. 2015. Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.5) University of Adelaide. Adelaide, Australia.

O'Brien, L.T., Bart, H.L. and D.M. Garcia. 2020. Why are there so few ethnic minorities in ecology and evolutionary biology? Challenges to inclusion and the role of sense of belonging. Social Psychology of Education 23: 449–477. https://doi.org/10.1007/s11218-019-09538-x

Pickrell, J. 2020. Scientists push against barriers to diversity in the field sciences. Science 374: 375. doi: 10.1126/science.caredit.abb6887

Revelle, W. 2019. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version =1.9.12.

Lenth, R. 2020. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.7. https://CRAN.R-project.org/package=emmeans

Saldaña, J. 2021. The coding manual for qualitative researchers. Los Angeles, CA: SAGE Publications.

Smith, D. 2004. Issues and trends in higher education biology fieldwork. Journal of Biological Education 39(1): 6-10.

Wilson, A. E., Pollock, J. L., Billick, I., Domingo, C., Fernandez-Figueroa, E. G., Nagy, E. S., ... and A. Summers, A. 2018. Assessing science training programs: Structured undergraduate research programs make a difference. BioScience 68(7): 529-534.

Supplementary Information

Supplementary 1. Field of study of FFC participants

Life Science
Ecology and Evolutionary Biology: 459
Ecology, Biology, Zoology: 22
Entomology: 34
Botany, Plant Biology, Plant systematics: 10
Natural Resources: 13
Neurobiology and Behavior: 2
Genetics:1
Molecular Cell Biology: 1
Immunology: 1
Physiology: 3

Other STEM
Applied Math: 15

Other fields
History: 2
History and Philosophy of Science and Technology, Science, Technology and Society: 9
International Agriculture & Rural Development: 3
Anthropology: 9
Civil Environmental Engineer: 1
International development: 1
Landscape Architecture: 1

NA: 15

Supplementary 2. Search protocol

Google Scholar:

1. From dropdown menu on left, select "Advanced Search"
2. **Author** - Return articles published by: "FirstinitialMiddleinitial [space] Lastname"
3. **Date** – Return articles dated between: Start Year – Last year PhD + 10
   a. Even though it says "between" this search will include the starting and ending years you provide
4. Click "Search"
5. On the left, unclick "citations" and "patents"
6. Scan results and tally relevant articles

Web of Science:

1. Click "Advance Search" on the main page near the search bar
2. **Author** – AU= Lastname [space] firstinitial*middleinitial*
   a. Don't put a space between the initials
   b. Last name comes first, then initials (opposite of google)
3. **Date** – Time span: Start Year – Last year PhD + 10
4. Click "Search"
5. If there are more than 1000 entries, you can refine by category (on the bar to your left)
   a. Web of Science will list the top 100 categories that the articles fall under. The top hits often include things that can be easily excluded ie – astrophysics or renaissance history. In general, leave things that could possibly be relevant, like "science history" or "medical entomology." Additionally, it is not worth eliminating things that only have a few hits. Unfortunately, searches will vary by author, and there are hundreds and hundreds of categories, so there isn't a list of things to exclude, just use your best judgement.
6. Scan results and tally relevant articles

Supplementary 3. Qualtrics Survey

Email:

Please complete the survey honestly and in one attempt - you cannot go back after you progress to the next page.

Block 1

1. Hello, the purpose of this survey is to gather feedback on independent research and field programs offered to graduate students at Cornell. We would love to hear about your impactful experiences as Cornell graduates, and where we can improve the program for future students.

2. What year during your graduate school experience did you start independently asking scientific questions (i.e. first year, second year, etc)? Give specific examples (3-5 sentences).

[Open end]

3. Please write about your early experiences in graduate school analyzing data and writing scientific papers. What experiences were most useful? Give specific examples (3-5 sentences).

[Open end]

4. Which of the following career pathways did you take after your PhD training? (select all that apply)

Faculty at a research-intensive institution

Faculty at a teaching-intensive institution

Academic non-tenure track, research career (e.g. postdoc, research associate, etc.)

Academic, non-research, non-teaching (e.g., curriculum development, program implementation,

grant writing, etc.)

Non-academic research career (e.g. industry, pharmaceutical, biotech science, govt or state

agency)

Non-academic, STEM, but non-research (e.g., science writing, high school science teacher, etc.)

Non-academic, non-STEM, non-research (e.g., lawyer, high school English teacher)

N/A (still in graduate school)

Other Please describe

5. Did you attend the Florida field course at Archbold during your graduate school experience?

Yes [Proceed to block 3]

No [Proceed to block2]

Block 2

6. You indicated that you did not attend the Florida field course at Archbold during your graduate school experience. Why not?

[Open end]

7. Did you attend other field courses during your graduate school experience?

If yes, please provide the name of the course(s) - Text

[Proceed to block Block 3]

If no, thank you for your time, your response has been recorded.

[Terminate]

Block 3

8. Please rate your level of agreement with the following statements:

9. The field course helped me ask more independent scientific questions

Completely disagree (1)

(2)

(3)

(4)

(5)

(6)

Completely agree (7)

10. The field course created a sense of community

Completely agree (1)

(2)

(3)

(4)

(5)

(6)

Completely disagree (7)

11. The field course helped me develop field technique proficiencies

Completely agree (1)

(2)

(3)

(4)

(5)

(6)

Completely disagree (7)

12. The field course taught me skills I applied in my PhD

Completely agree (1)

(2)

(3)

(4)

(5)

(6)

Completely disagree (7)

13. The field course taught me skills I apply in my career

Completely agree (1)

(2)

(3)

(4)

(5)

(6)

Completely disagree (7)

N/A Unsure [off scale]

14. Based on your experience, when is the best time to participate in the field program?

Select all that apply.

Year 1

Year 2

Year 3

Year 4

Year 5

Year 6

Any/no preference

Block 4

15. Thank you for your responses so far, we are almost done.

For the following questions please write 3-5 short sentences reflecting on your field course experience:

16. **What would you say to a graduate who was considering the program?**

    [Open end]

17. **What was the most impactful learning or skill you took away from the program?**

    [Open end]

18. **What would you change or improve about the program?**

    [Open end]

19. **Is there anything else you'd like us to know?**

    [Open end]

Thank you for your time, your response has been recorded.

Supplementary 4. FFC Coding Rubric

Supplemental Table S4.1. Coding rubric for early experiences in graduate school that were beneficial to writing and data analysis.

Theme	Description of Theme	Examples
Student Led Environment	Meetings or sessions where peers provide feedback or share skills.	"A course in R programming taught by another student"
Mentor Feedback	Feedback from a mentor-figure, either at the student's own institution or another institution.	"I had an excellent mentor in terms of paper writing and the push to write papers. This mentorship and guidance was invaluable. "
Statistics Consulting Unit	The Statistics Consulting Unit is explicitly mentioned.	"Stats consulting unit (free in-person help) helped me a lot."
Graduate Classes	Specific courses or classes that were useful.	"I took two semesters of biostats, which were really helpful as I started to analyze my own data."
Seminars	Conferences, workshops, or presentations the student attended and found helpful for developing writing and/or data analysis skills.	"the seminar series (both the EEOB seminar and the BGC seminar) were excellent"
Required Task/ Independence	Required tasks, or independently conducted activities that were helpful for developing writing and/or data analysis skills. This code was not assigned if used in the context of another listed theme.	"Writing papers with a great deal of independence was useful."
Field Course	Course attended by the student and referred to as a "field course."	"I joined the Florida Field Course early in my grad experience and this was useful"
Grant Application	Grant application is explicitly mentioned as a useful way to develop writing and/or data analysis skills.	"It was writing for real things (like grant proposals and papers) that were most valuable"

Collaboration	Mentions "working with others", or "collaboration" in some way. Includes students, instructors, and other faculty specifically working on a project.	"Most useful was collaborations, with faculty and with other graduate students." "Also, co-authoring papers with other students on topics outside my immediate research area helped me to learn about writing papers."

Supplemental Table S4.2. Coding rubric for the most impactful learning or skill students took away from the field course program

Theme	Description of theme	Examples
Carrying out research	Mentions the opportunity to ask a research question, develop an experiment, analyze data, or engage in the scientific process. Can include time management skills.	"Thinking more about how to analyze complex ecological data."
Mentor/peer feedback	Reports "criticism" or "feedback" from the scientific community during the program.	"Building confidence in myself and in my work as a scientist, learning how to stand up for myself and change with constructive criticism at the same time."
Instruction/presentation skills	States that "presenting" was helpful or that they learned skills that could aid in future teaching or science communication.	"Working with first year students, I realized how much they had to learn. It was a little frustrating but probably good preparation for teaching."
Scientific writing/grant writing	Mentions writing about science was an important learned skill for papers or grant proposals.	"Grit and a thicker skin (learned via winning funding, designing and doing fieldwork, and writing and revising and revising some more)"
Learning/observing from nature	Being in the environment and making observations about nature.	"Focus on field-based ecology, close observation, perfect environment for someone interested in plant-animal interactions"

Thinking broadly about science/ curiosity/discovery	Exposure to other scientific research areas and topics, being open minded.	"The ability to think broadly about ecology and evolution."

Supplementary 5. Number of students per cohort

Cohort	**Count**
1	2
2	103
3	106
4	83
5	89
6	98
7	86
8	30
9	NA
NA	6

Appendix 4. Why students struggle in undergraduate biology: sources and solutions (In revision, formatted for *CBE-Life Sciences Education*).

Contribution: Performed qualitative data analysis; Co-wrote the paper; Revised and edited drafts of the paper.

Manuscript Title: Why students struggle in undergraduate biology: sources and solutions

Running Title: Student struggle in biological sciences

Author list:
Claire B. Tracy[1*], Emily P. Driessen[1*], Abby E. Beatty[1], Todd Lamb[1], Jenna E. Pruett[1,2], Jake D. Botello[1], Cara Brittain[1], Ísada Claudio Ford[1,3], Chloe C. Josefson[4], Randy L. Klabacka[1], Tyler Smith[5], Ariel Steele[6], Min Zhong[1], Scott Bowling[1], Lucinda Dixon[7], and Cissy J. Ballen[1]

[1]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA
[2]Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO 80309, USA
[3]Department of Natural Science, University of Puerto Rico, Rio Piedras, Puerto Rico
[4]Department of Animal, Veterinary, and Food Sciences, University of Idaho, Moscow, ID 83844, USA
[5]Department of Geosciences, Auburn University, Auburn, AL 36849, USA
[6]Department of Educational Foundations, Leadership, and Technology, Auburn University, Auburn, AL 36849, USA
[7]Lincoln Memorial University, Harrogate, TN 37752, USA
*equal contributions

Corresponding author: Claire Tracy (cbt0022@auburn.edu)

Keywords: performance, undergraduate, introductory, classroom, course deficit model

Character Number (Inc. Spaces): 72,831

Manuscript for consideration as an Article in *CBE-Life Sciences Education*

ABSTRACT

Students' perceptions of challenges in biology influence performance outcomes, experiences, and persistence in science. Identifying sources of student struggle can assist efforts to support students as they overcome challenges in their undergraduate education. In this study we characterized student experiences of struggle by (1) quantifying what external factors relate to perceptions of encountering and overcoming struggle in introductory biology, and (2) identifying to what students attribute their struggle in biology. We found a significant effect of course, instructor, and incoming preparation on student struggle, where students with lower incoming preparation were more likely to report struggle and the inability to overcome struggle. We also observed significant differences in performance outcomes between students who did and did not encounter struggle and between students who did and did not overcome their struggle. Using thematic and open coding we categorized student responses outlining causes of struggle and further categorized these as internally or externally attributed factors. External sources (i.e., Prior Biology, COVID-19, External Resources, Classroom Factors) were more commonly cited as the reason(s) students did or did not struggle. We conclude with recommendations for instructors, highlighting equitable teaching strategies and practices.

INTRODUCTION

Many students who enter higher education with the intent to pursue a career in the biological or biomedical sciences abandon this goal because they struggle in introductory 'gatekeeper' science courses (Gainen, 1995; Gasiewski et al., 2012). Promoting student retention in STEM is essential to our national efforts to produce graduates that meet the growing need for a trained and diverse workforce (President's Council of Advisors on Science and Technology, 2012). This goal may be addressed through mitigation of student struggle in STEM courses since struggle can significantly undermine students' academic abilities and performance (Batz et al., 2015; England et al., 2017a). While the term is ill-defined in the literature, internal and external factors can contribute to students' experiences of struggle.

Internal Factors

Academic success and struggle are influenced by students' content knowledge as well as a host of internal, or affective, factors within the student's control (Austin et al., 2018; Ballen et al., 2017; Cooper & Brownell, 2020). One influential internal factor is the mindset of a student when encountering struggle (Yeager et al., 2019), and this contributes to performance outcomes. Additionally, previous work shows that students' beliefs about the degree to which intelligence is a stable trait is an influential factor impacting student struggle (Dweck, 1999; Limeri et al., 2020). Specifically, Dweck (1999) found that students who perceived intelligence as an unchangeable trait, or an innate ability, were more likely to interpret struggle as an indication that they were not intellectually capable of success. However, we know student abilities are not fixed because individuals' mindsets develop and change throughout their lives (Aronson et al., 2002; Yeager et al., 2019). To address how mindset changes over time and how it can be influenced by STEM coursework, Limeri et al. (2020) used latent growth modeling to demonstrate that students who reported they had struggled in a course also increasingly viewed intelligence as an unchangeable trait over a semester.

Other internal factors that may cause students to perceive struggle in relation to their undergraduate biology education include study habits, motivation, self-determination, and grit (passion and perseverance), as well as sense of belonging in STEM (Cromley et al., 2016; Dyrberg & Holmegaard, 2019; Flanagan & Einarson, 2017; Johnson et al., 2007; Meaders et al., 2020; Ryan & Deci, 2017, 2020; Sheshadri et al., 2019; Sithole et al., 2017; Wilton et al., 2019; Ye et al., 2016).

External Factors

Sources of struggle can also be due to external factors which are outside of the student's control. External factors that may cause students to perceive struggle include course format (e.g., regular use of undisrupted lecture or unstructured group work), reliance on high-stakes exams to evaluate students, and large class sizes, particularly in introductory courses (Armbruster et al., 2009; Ballen et al., 2017; Corkin et al., 2017; Freeman et al., 2014; Gasiewski et al., 2012; Scott et al., 2017). Two other external factors contributing to students' perceived struggle are previous educational experiences and access to resources, which play a central role in student incoming academic preparation for higher education and overall success in entry level college courses (Freeman et al., 2007; Salehi et al., 2019, 2020).

We include incoming preparation as an *external* factor to emphasize its reflection of the opportunity gap in primary and secondary education. For example, access to high-quality curricular materials, evidence-based instruction, and technology all contribute to gaps in academic preparation, which are subsequently reflected in student grades in introductory courses that fail to provide students with equal opportunities to demonstrate proficiency in course content knowledge (Salehi et al., 2019, 2020). Specifically, several analyses of performance outcomes in STEM at multiple institutions revealed that differences in student performance appeared to be due to differences in SAT, ACT, and pre-semester concept inventory scores (Salehi et al., 2019, 2020). Thus, we may predict a strong relationship between measures of incoming preparation, course performance, and perceptions of struggle. To our knowledge, this relationship has not been studied in undergraduate biology. While incoming preparation is often linked to performance outcomes, it has also been well demonstrated that some instructional practices and classroom factors can expand or reduce (or eliminate!) incoming gaps in academic preparation. For example, instructors who use evidence-based and equitable teaching strategies can decrease performance gaps in undergraduate STEM courses (Ballen et al., 2017; Cotner & Ballen, 2017; Freeman et al., 2007; Salehi et al., 2020; E. J. Theobald et al., 2020).

Finally, a unique potential source of external struggle for students in Spring 2020 was the emergency transition to remote learning amidst the COVID-19 pandemic. At its onset, higher education institutions required students shelter in place for an indefinite period of time, while functionally cutting students off from essential learning resources on campus. The transition to emergency remote learning dramatically impacted how students prepared for and took exams (Beatty et al., 2022; Driessen, Beatty, et al., 2020), participated in the classroom setting (Ali et al., 2020; Wester et al., 2021), and developed social relationships with both peers and instructors (Smoyer et al., 2020; Supriya et al., 2021; Wut & Xu, 2021). The transition to remote learning also created practical challenges for faculty, many of whom resorted to asynchronous online videos or a combination of asynchronous and synchronous work, with little to no instructor-student interaction (Supriya et al., 2021; Wolinsky, 2021). While pedagogical decisions were the result of dire circumstances, the impacts of those decisions were still consequential to student learning and potentially, perceptions of struggle in the course.

Research Questions

The emergency transition to remote learning amidst the COVID-19 pandemic created an opportunity to study ways in which students attribute their success and struggle in biology during one ‘typical’ semester and one in which a significant disruption took place. Importantly, we did not define struggle for the students, but rather used their open-ended responses and course performance to better understand their view/interpretation of struggle while addressing the following research questions:

Part I: (a) What factors influence struggle in introductory biology? (b) How do performance outcomes correlate with students encountering and overcoming struggle in introductory biology?

Part II: (a) To what do students attribute their struggle (or lack of struggle) in introductory biology? (b) What sources of struggle were students most likely to overcome?

METHODS

Courses and Instructors

We examined data from a survey of 965 students across three different Introductory Biology classes (hereafter Courses 1-3 in order of sequence) with 6 different sections during the Fall 2019 and Spring 2020 semesters (Table 1). We consider these courses introductory because they are taken by mostly first- and second-year undergraduate students as pre-requisites for upper-level Biology courses. Participating courses were part of the Biology department at a large, primarily white public research university in the southeastern region of the United States. Enrollment for each lecture section ranged from 159-376 students. The courses surveyed were taught by five instructors, whom varied in terms of teaching experience and demographics, but all instructors utilized active learning components. Following Driessen et al. (2020), we define active learning as “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”. The specific strategies instructors employed varied, ranging from flipped classroom to group work with iClicker questions. Instructor teaching experience varied between 0-18 years.

Data Collection

Data collection and research was approved by Auburn’s Institutional Review Board #18-349 EP 1811. Prior to data collection, we recruited instructors via email and other personal communications. From the consenting instructors’ classes, we collected data from their enrolled undergraduate students through the use of an online survey created in Qualtrics (Table S1). We posted the survey on the Canvas page for each of the participating classes during the last week of the semester for the Fall 2019 and Spring 2020 semesters. Each instructor made their students aware of this optional survey, some of which provided bonus points for participating, and students either did not take the survey, took the survey and consented as a participant, or took the survey and did not consent to participate. We used only the data from the consenting participants. At the end of the semester, we additionally collected institutional information for all of the students enrolled in the six classes participating in this study. This information included student grades, ACT and/or SAT scores, high school GPA, college GPA, sex, and race/ethnicity as obtained from the Office of Institutional Research.

The survey instrument opened with an information letter, detailing the purpose of the study to participants, asking for students to consent or not consent, prior to accessing the rest of the survey. The remainder of the survey collected demographic information (i.e., gender identity and race/ethnicity) and open-ended responses to the following questions: (1) “Did you encounter struggle in introductory biology this semester?” and, if so, (2) “Were you able to overcome struggle that you encountered?” (Table S1). The survey questions used in this research were previously developed and implemented to 875 students in an Organic Chemistry II course as detailed in Limeri et al. (2020). Students followed logic questions, an advanced survey option that allows the creator to include ‘rules’ within the survey (L. Limeri, personal communication, September 8, 2020). By implementing logic, students advanced to further questioning that was specific to their earlier responses (Figure 1). We operationalized perceived struggle using two “Yes/No” questions, ensuring our interpretations of student responses to these questions accurately reflected their experiences by presenting our interpretation to students and asking them to comment on it in a constructed-response question. For example, students who answered "No" to the first question were then asked, "Your response suggests that you did not encounter struggle in [biology course] this semester. Please explain why this is or is not an accurate description of your experience." Written responses to these questions were reviewed during the qualitative coding process (see Part II: Qualitative Analysis), and any written responses that contradicted their binary Yes/No response were removed from analysis.

Of the 1453 students enrolled, we received 965 total survey responses with 130 students taking the survey both in Fall 2019 and Spring 2020 for different courses and 3 students taking the survey twice in the same semester for different courses. In the total dataset of 965, 5 responses were removed as their open-ended responses contradicted their binary response (e.g., a student selected No, indicating they did not experience struggle, but they then went on to write a response suggesting they did struggle). The remaining 960 responses were used in our quantitative analysis. Of the 960 responses from the quantitative dataset, only 745 responses contained complete open-ended qualitative responses. Of the 745 responses, 38 were discarded because their open-ended responses could not be confidently coded into one of the codes. This left us with 707 student responses for our qualitative analysis. A portion of the 707 entries with open-ended qualitative responses included duplicate responses by students in different courses (82 students) (see Limitations section for more information). For a summary of the demographics of the enrolled and participating students from the accumulative six sections, see Table S2.

Part I: Quantitative Analysis

1A. What factors influence struggle in introductory biology?

We first were interested in what factors were correlated with struggle in Introductory Biology courses. We focused on how incoming preparation (i.e., access to resources), instructor, and course, which are strongly linked to overall course performance, are correlated with struggle.

We used a principal component analysis (PCA) to collapse two measures of incoming preparation into one variable. Incoming preparation measures include a standardized test score (ACT scores or SAT scores converted to ACT scores) and high school GPA. SAT scores were converted to ACT scores following the ACT/SAT concordance tables provided by ACT.org. In the PCA, PC1 explained 76% of the variance in the dataset and was extracted for use as a single measure of incoming preparation.

To determine the relationship between incoming preparation, instructor, and course on student struggle, we used model selection on generalized linear mixed-effects models (GLMM) to determine the best fit structure of both fixed effects and random effects. The fixed effects we used in model selection included our principal component of incoming preparation (PC1), Instructor, and Course. In the case where there were two instructors for a single section (alternating instructors), we treated both instructors as a single instructor, as the students remained the same for that course, and neither instructor taught independently in courses surveyed in this study. The random effects we included in model selection to account for variation in the dataset include Student ID (to account for having duplicate entries from 130 individuals in our dataset), Student Year Rank (first-year, second-year, third-year, fourth-year), Section (to account for potential differences between sections for a single instructor), and Semester (to account for the expected variation that existed between a pre-COVID semester and a semester with COVID and the emergency transition to remote learning. Following best practices outlined by Theobald (2018), we determined the best fixed effects structure first without including random effects, and subsequently determined the best random effects structure while holding the fixed effects constant. The best fit model was selected using Akaike's Information Criterion (AIC).

1B. How do performance outcomes correlate with students encountering and overcoming struggle in introductory biology?

To further understand how students are impacted by struggle, we were interested in understanding how student experiences of struggle impacted overall performance outcomes in the course when accounting for factors that are known to influence performance such as Incoming Preparation, Instructor, and Course. We used linear mixed-effects modelling to determine the best model predicting final score in the course. Fixed effects included in model selection were incoming preparation and encountering or overcoming struggle. While we were not directly interested in the relationship between Incoming Preparation and performance, we included this factor to account for any relationship that exists. We included Incoming Preparation as a fixed effect as it is a continuous numerical variable and as such cannot be treated as a random factor. We determined the best-fit model using AIC score comparison as outlined above in part 1A. Random effects included in model selection were Student ID, Student Year Rank, Instructor, Course, Section, and Semester. Further explanation on why each factor was included in model selection can be found above in part 1A.

Part II: Qualitative analysis

After downloading the answers to the open-response question (Figure 1), nine of the co-authors (AEB, JP, TL, JDB, CB, ICF, CCJ, TS, CJB) individually reviewed a set of 40 student responses to the open-ended question and generated codes using inductive coding (deduced codes from data rather than creating codes a priori; Saldaña, 2009) and qualitative content analysis (i.e., a tool used to determine the presence and frequency of certain codes within the open-ended responses; Morgan, 1993). They also took detailed analytic notes at that time (Birks & Mills, 2015). They then met together to compare their codes and revise the rubric. The researchers used constant comparison methods to ensure quotes within a code were not too different from each other to warrant the creation of a new code (Glesne & Peshkin, 1992). This process was repeated until the group was confident with their rubric, developing one unified coding rubric with detailed descriptions and examples. The final ten codes in the rubric were: (1) Prior STEM (outside of biology), (2) Prior Biology, (3) COVID-19, (4) External Resources, (5) Classroom Factors, (6) Study Habits, (7) Innate Ability, (8) Time Management, (9) Preference for Biology, and (10) Anxiety. During the code creation process, we noticed students referenced the code Classroom Factors frequently, so one of the co-authors (TL) further broke down the Classroom Factors code into six subcodes, (1) Content; (2) Exams; (3) Format; (4) Group Work; (5) Instructor; and (6) Workload. Sources of struggle and reasons student did not encounter struggle were seamlessly categorized in the same way, and always represented two sides of the same coin (i.e., two separate parts of the same category). For example, a student may describe their struggle because they were unfamiliar with biology or STEM content, or they did *not* struggle because they had a strong background in biology or STEM; or a student may have mentioned that they did struggle because of the large workload, or they did *not* struggle because they had a manageable workload.

After the ten main codes were established, the same nine co-authors grouped the codes into one of two categories: internally or externally attributed struggle. Codes 1-5 were considered to be external attributes (i.e., outside of the student's control) while codes 6-10 were categorized as internal (i.e., within the student's control). Student responses to their reported struggle and ability to overcome reported struggle were assigned to all appropriate codes, meaning that a single response could fit in more than one category.

Once the codes, sub-codes, and categories were established, three of the co-authors (EPD, CBT, TL) completed the coding independently in three sections: (1) students who did not struggle, (2) students who did struggle but did not overcome, and (3) students who did struggle but did overcome. For each section, each of the three coders independently coded responses in "blocks" ranging from 40 to 169 responses and then collaboratively coded to consensus. This resulted in a total of seven different blocks, with an average initial percentage agreement of 69.5% for the main codes and 69.8% for the Classroom Factors subcodes. We calculated percent agreement by dividing the total number of codes agreed on by the number of codes agreed on plus the number of codes not agreed upon. We calculated percentages for each code by dividing the total number of responses assigned for each code by the total number of student responses for each category.

RESULTS

Descriptive statistics

After downloading the Qualtrics survey results, we noted the percentage of students who did or did not encounter struggle, and of those who experiences struggle we reported those who did and did not overcome it. Of the 960 responses, 253 (26%) reported that they did not encounter struggle (N=63 Fall, N=190 Spring) and 707 (74%) reported that they did encounter struggle (N=382 Fall, N=325 Spring). Of the 707 that encountered struggle, 190 (27%) reported they did not overcome their struggle (N=101 Fall, N=89 Spring), and 515 (73%) reported that they did could overcome their struggle (N=280 Fall, N=235 Spring), and 2 didn't respond to whether they overcame or not.

Of the 707 responses that included complete open-ended qualitative responses, 560 (79%) reported encountering struggle (N=307 Fall, N=253 Spring) and 147 (21%) reported they did not encounter struggle (N=47 Fall, N=100 Spring). Of the 560 that struggled, 412 (74%) report overcoming their struggle (N=228 Fall, N=184 Spring) while 148 (26%) report not being able to overcome their struggle (N=79 Fall, N=69 Spring) (Figure S1).

Part I: Quantitative Analysis

1A. What factors influence struggle in introductory biology?

The best fit model for predicting if students encountered or overcame struggle was determined by identifying the model with the lowest AIC score (Table S3 & S4). The best fit model for encountering struggle included the interaction between our measure of incoming preparation (i.e., the results of a principal component analysis described in methods; hereafter PC1) and Course. Student Year Rank was included as a random effect. During model selection, we also included Instructor, Student, Section and Semester as possible additional factors, but the best fit model did not include these factors as random or fixed effects (see Methods for more detail on model selection). The best fit model for overcoming struggle included PC1 and Instructor as fixed effects, but not an interaction between the two, and included student as a random factor.

*Encounter Struggle ~ PC1 * Course + (1 | Student Year Rank)*
Overcome Struggle ~ PC1 + Instructor + (1 | Student ID)

Incoming preparation had a significant effect on encountering and overcoming struggle (Tables S5 & S6). The model showed that students with higher measures of incoming preparation were less likely to encounter struggle and more likely to overcome struggle if they did encounter it. Specifically, with each point increase in incoming preparation, students were 0.66 times less likely to encounter struggle and 0.25 times more likely to overcome their struggle.

Course and Instructor also both had significant effects, where Course was significantly correlated with encountering struggle and had an interaction with Incoming Preparation (Table S5, Figure S2), and Instructor was significantly correlated with students overcoming struggle (S6, Figure S2). Specifically, students in Course 2 and Course 3 were less likely to encounter struggle (1.1 times and 1.4 times less likely, respectively) than students in Course 1 (Table S5, Figure S2). We contextualize these findings by pointing out that Course 1 is the first Introductory Biology course that most students take their first semester of college, and is a prerequisite for Course 2 and Course 3. Additionally, Student Rank (first-year, second-year, third-year, fourth-year) was included in the best fit model for encountering struggle as a random factor. The model accounting for both Student Rank and Student ID had a delta AIC of 2, indicating that they were identical in fit, and as such we presented the most parsimonious model.

1B. How do performance outcomes correlate with students encountering and overcoming struggle in introductory biology?

We next wanted to determine if students' experiences of struggle correlated with their final score in the course, controlling for incoming preparation. The best fit model was determined by comparing AIC (Table S7) and included two fixed effects (Encountering Struggle and Incoming Preparation (PC1)), and two random effects (Instructor and Student ID):

Final.Score ~ Encounter_Struggle + PC1 + (1 | Instructor) + (1 | Student ID)

We found a significant effect of encountering struggle on performance in the course when also accounting for variation due to Incoming Preparation (PC1) (Table S8, Figure 2). Students who encountered struggle scored 3.7 points lower on their final grades than students who did not struggle.

We then examined the effect of overcoming struggle on final score while accounting for effects of Incoming Preparation, as well as Instructor and Student ID as random effects. The best fit model was found using AIC comparison to be as follows (Table S9):

Final.Score ~ Overcome_Struggle + PC1 + (1 | Instructor) + (1 | Student ID)

There was a significant effect of overcoming struggle on performance in the course when accounting for variation due to Incoming Preparation (PC1) (Table S10, Figure 2). Students who overcame struggle had final grades 3.6 points higher than students who did not overcome their struggle.

Overall, students who reported not encountering struggle performed slightly better when accounting for incoming preparation than students who reported encountering struggle; students who reported encountering and overcoming struggle performed slightly better than students who encountered but could not overcome struggle when accounting for Incoming Preparation (Figure 2). **This indicates that students who encountered struggle and those who could not overcome struggle received lower grades than students who did *not* encounter struggle and those who reported overcoming struggle.**

Part II: Qualitative analysis

For ease of interpretation of the qualitative results, we developed a qualitative code key for the ten codes, as binned into the broader two categories of external and internal attributes, complete with code explanations and both a positive and negative example of student use (Figure 3). For example, if a student said they did not struggle or overcame their struggle because of something, then that would be a *positive* example of the category. However, if a student said they struggled, their response would be a *negative* example of the category. Additionally, we provided a breakdown of the subcodes of Classroom Factors with explanations of the codes and both positive and negative examples (Figure 4). We recommend referral to Figure 3 and Figure 4 to understand the codes and examples. We present results from both semesters (e.g., Fall 2019 and Spring 2020) separately since there were distinct circumstantial differences between the two (e.g., different instructors, different students, different semesters, and the pivot in Spring 2020 to emergency remote learning due to the COVID-19 pandemic).

2A. To what do students attribute their struggle (or lack of struggle) in introductory biology?

When we asked students to elaborate on their experiences, we found explanations differed based on whether students did or did not struggle (Figure 5A). Across both semesters, we found students who did not face struggle largely attributed their lack of struggle to Classroom Factors (31% Fall, 34% Spring), Study Habits (29% Fall, 27% Spring), and Prior Biology (29% Fall, 16% Spring). One response that typifies a positive reference to Classroom Factors is, “The information was fun and interesting to learn and the teachers provided multiple opportunities to retain the information, so this allowed me to avoid struggle” (Figure 3).

Among students who reported facing struggle, the most often cited reasons for struggle were Classroom Factors (65.6% Fall, 59.7% Spring) and Study Habits (23.5% Fall, 21.8% Spring). A response that typifies a negative reference to Classroom Factors is, “I struggled with assignments that were not announced to be completed and the failure to announce what to review before classes” (Figure 3). However, the Classroom Factors code had a wide variety of responses which led to a more detailed breakdown of this code into subcodes (Figure 4). When comparing students who did not struggle with those who did struggle, we observed that students who did struggle mentioned Classroom Factors more, Prior Biology less, Time Management more, and Study Habits less (Figure 5A).

We found that students – both those who did and did not struggle – largely referred to external factors (e.g., Prior STEM, Prior Biology, COVID-19, or Classroom Factors) more than internal factors (e.g., Anxiety, Innate Ability, Preference for Biology, Study Habits, or Time Management; Figure 6). However, students who did not struggle referred to internal reasons (26% Fall, 32% Spring) more often than students who did struggle (18% Fall, 19% Spring).

Classroom Factors Reference Breakdown. When comparing referenced classroom factors between students who did and did not struggle (Figure 5B), major differences included students who struggled more often cited Workload, Exams, and Format as sources of their struggle.

In contrast, students who did not struggle more often cited the Instructor, Group Work, and Content as a reason for their lack of struggle. The most cited reasons for not struggling (when referencing Classroom Factors) were the Instructor (30% Fall, 52.6% Spring), the Format (30% Fall, 7% Spring), and the Content (25% Fall, 22.8% Spring). For example, students who cited Instructor as a reason they did not struggle (i.e., a positive reason) often indicated a response similar to "I felt my professor prepared us and I worked very hard outside of class" (Figure 4). The most often cited reasons for struggle when referencing Classroom Factors were the Format (34.8% Fall, 30% Spring), Exams (20.6% Fall, 27.3% Spring), and Content (16.1% Fall, 21.8% Spring) (Figure 5B). One example of a response where Format was a source of struggle (i.e., a negative reason) was "It was very difficult for me to learn in a classroom that was flipped."

2B. What sources of struggle were students most likely to overcome?

Reasons for struggle largely varied depending on whether the student could overcome or not (Figure 5C). For example, students who did not overcome their struggle most often cited Classroom Factors as their source of struggle (76% Fall, 68% Spring). Students who did overcome their struggle most often cited Classroom Factors (62% Fall, 57% Spring) and Study Habits (28% Fall, 27% Spring) as their source of struggle.

When examining sources of struggle as external factors (e.g., Prior STEM, Prior Biology, COVID-19, or Classroom Factors) and internal factors (e.g., e.g., Anxiety, Innate Ability, Preference for Biology, Study Habits, or Time Management), we found that students who overcame their struggle cited internal factors as their reason for struggle (22% Fall, 24% Spring) more than students who did not overcome (9% Fall, 9% Spring; Figure 6).

Classroom Factors Reference Breakdown. In both semesters, there were not major differences in sources of struggle attributed to the classroom among students who did or did not overcome struggle (Figure 5D). For example, the most cited sources of struggle for **students who did not overcome** were the Format (38.9% Fall, 36% Spring), Exams (22.2% Fall, 25.3% Spring), and Instructor (21.1% Fall, 12% Spring). Similarly, the most cited sources of struggle for **students who did overcome** were Format (32.8% Fall, 26.9% Spring), Exams (19.8% Fall, 28.3% Spring), and Content (21.5% Fall, 25.5% Spring). The largest differences between students who did and did not overcome their struggle were that students who did not overcome cited Format and Instructor more, while students who did overcome their struggle cited Content more.

DISCUSSION

Part I: Quantitative Analysis

What factors impact struggle and how does struggle impact overall performance?

The linear models show a significant effect of incoming preparation on both encountering and overcoming struggle. We also found a significant effect of encountering and overcoming struggle on performance, even when accounting for incoming preparation.

Incoming preparation was significantly correlated with encountering and overcoming struggle. This was not surprising because of the established relationship between incoming preparation and STEM course performance (Salehi et al., 2019, 2020) and because performance is likely one way that students gauge perceptions of struggle. Additionally, while it appears at face value that struggle explains performance, students' perceptions of struggle are likely driven by their performance in the course (though with our current data, we cannot disentangle directional impacts). Nevertheless, given our measure of incoming preparation relates to precollege educational resource availability, and that struggle was experienced disproportionately by students with lower incoming preparation, our results provide more evidence that 'gateway' courses differentially impact subsets of students. Future areas of research should explore effective teaching strategies to help students overcome challenges in introductory biology courses, as well as strategies students use to cope with obstacles and struggle.

Our results also support previous work showing student perceptions of their capacity to succeed within a discipline are driven by performance outcomes in STEM (Seymour & Hunter, 2019). Some students will underestimate their abilities based on grades compared to similarly performing students (Marshman et al., 2018). Mentions of performance arose multiple times in students' open-ended responses, which would often express the sentiment that they felt they knew the material but performed poorly on the exams. Students also often cited poor grades on exams when asked why they struggled in the course. Many works have demonstrated the importance of grades to undergraduate students (DeFeo et al., 2021; Lewis, Williams, Sohn, & Loy, 2017; Sabot & Wakeman-Linn, 1991), which likely carries over into their perceptions of struggle. One student responded to the open-ended prompt saying "I made a C on the first test and did very well on everything else. Completed all assignments and did well on the next tests and still looks like I'll finish with a B", and while to many this would sound like a student who encountered struggle and overcame it, this student indicated that they encountered struggle and did not overcome it, demonstrating a strong link between grades and perceptions of struggle. Other students viewed struggle differently, as one student who expressed a similar sentiment of struggling on the first test but doing better on the rest of the exams categorized themself as not encountering struggle.

Our results also support the idea that STEM classes appraise students' abilities to understand science, and this appraisal is related to how well their previous schooling prepared them for tertiary education. This puts capable but academically underprepared students at a serious disadvantage (Salehi et al., 2019, 2020). Intentional or not, the practice of using STEM courses to weed out students seeking degrees in STEM is problematic because it hinders efforts to attract students who have been historically excluded in those fields, such as underrepresented minority students, first generation college students, and women (Mervis, 2011).

In order to address gaps in incoming preparation, it is important to tailor teaching in introductory courses to current levels of incoming preparation, and provide greater resources that can aid in bridging any existing gaps in incoming preparation such as supplemental instruction programs or peer-tutoring study groups (Batz et al., 2015; Meaders et al., 2020; Salehi et al., 2019). Another commonly discussed alternative that may help make introductory biology courses more accessible across different preparation levels is a larger emphasis on low-stakes formative assessments. Shifting focus to lower-stakes assessments has been shown to provide a mechanism to assess student knowledge without significant performance gaps that are often apparent in high-stakes assessments among genders or students historically excluded in science because of their ethnicity or race (Cotner & Ballen, 2017; Sambell & Hubbard, 2004).

In addition to incoming preparation being correlated with struggle, there was also a significant correlation between course and encountering struggle. The course in which students were more likely to encounter struggle was Course 1, which is the first biology course that most biology majors take in college, often in their very first semester. This sentiment was prevalent in students' open-ended responses, where many students identified their sources of struggle being related to study habits, not knowing how or what to study for exams, and adjusting to college courses in general (see Part II for further discussion on study habits). Instructors can help mitigate struggle experienced by students just starting college by spending class time early in the semester discussing study habits, time use both in and out of class, and by setting clear expectations for the course, particularly in introductory courses.

Our results also suggest that instructors can influence students overcoming struggle. We do not break down our results by instructor or teaching practices, and thus cannot identify specific teaching practices or other classroom factors that differ between instructors. However, in Seymour and Hunter (2019), interviews revealed students characterized good teachers and teaching through several characteristics, including teachers who (1) showed concern for student learning, (2) engaged students during class, (3), provided course structure which was organized and coherent, (4) used interactive reaching methods, and (5) made connections between the material and the real world.

Part II: Qualitative Analysis

Our results demonstrated that students' perceived sources of struggle included a number of factors such as experience in Prior STEM or Prior Biology courses, COVID-19 (only during Spring 2020), External Resources, Classroom Factors, Study Habits, Innate Ability, Time Management, Preference for Biology, and/or Anxiety. As Spring 2020 was a tumultuous semester for many due to the emergency transition to remote learning secondary to the onset of COVID-19 pandemic mid-semester, it was not necessarily reflective of a normal or repeatable semester. Further, Spring 2020 results were very similar to Fall 2019 results, aside from the novel code: COVID-19. For these reasons, this Discussion will focus on the results from Fall 2019.

We will first focus on the top three most mentioned codes: (1) Classroom Factors, (2) Study Habits, and (3) Prior Biology. Then, we will follow with a discussion of bigger picture trends in codes grouped as either internal or external sources of struggle.

Classroom Factors

Overall, both students who faced struggle and those who did not face struggle elaborated by citing Classroom Factors, demonstrating how specific factors can be advantageous to some students while disadvantaging others. Within this category, the two most mentioned subcodes were Instructor and Format.

First, students commonly cited Instructor as a reason they did not face struggle (e.g., "I felt my professor prepared us, and I worked very hard outside of class"), while a much smaller percentage cited the Instructor as a reason why they did struggle (e.g., "the professor just didn't click with me"). However, sources of struggle that students were *least* likely to overcome were also Classroom Factors, including instructor and format. Previous literature highlights the importance of the instructor, demonstrating that STEM professors with fixed mindsets caused students to feel less of a sense of belonging and more stereotyped based on their gender as opposed to those instructors with growth mindsets. This ultimately negatively impacted women students and their performance in the course (Canning et al., 2021). Canning et al. (2019) showed professors' beliefs about the fixedness of ability were associated with racial performance gaps that were twice as large as gaps in courses taught by more growth mindset faculty. Similarly, student-reported trust of the instructor corresponded to their final grade (Cavanagh et al., 2018). Seidel et al. (2015) created an "Instructor Talk framework" to assist instructors in reflecting on the learning environments they create through non-content language in classrooms. When utilizing their framework with two different instructors, Seidel et al. (2015) discovered specific subcodes of instructor talk may play a critical role in constructing inclusive environments, allowing the potential for students to overcome stereotype threat (i.e., boosting self-efficacy, revealing secrets to success, and promoting diversity in science). One example of instructor talk focusing on promoting diversity in science from the Seidel et al. (2015) study reads:

> We absolutely know, we have lots of stories that say the kinds of people who do science affect the kinds of questions that get asked, affect the kinds of data that gets acknowledged, and the kind of data that gets ignored. So, that's why it's really important to have a diverse group of people doing science (p. 6).

We recommend instructors create an inclusive learning environment that signals safety for all students in the classroom by utilizing the framework detailed in Seidel et al. (2015) to monitor and improve the message they convey to their students about the classroom and themselves as individuals.

Second, both students who did struggle and those who did not struggle mentioned Format in their responses. Additionally, of those students who cited Format as a source of struggle, less of these students overcame this struggle than those that did. This finding may be explained by a preference for or against active learning, the dominant pedagogy deployed across classrooms in the current study. This is supported by previous literature that demonstrated undergraduate students learn more in classrooms that use active learning than those that use traditional lecture strategies, even though they perceive they learn less (Deslauriers et al., 2019). This hypothesis is well supported by the following quote obtained from a student facing struggle in the Fall 2019 semester: "I, along with many of my peers, did not like the "flipped classroom" curriculum. It was said that many students learn better this way, but the average test scores do not support this. A regular lecture-based classroom is a better way for students to learn. Teaching ourselves provides a workload that isn't sustainable for the average college student." Student comments along these lines were common during the Fall 2019 semester.

Alternatively, this finding could demonstrate that the given format chosen by a professor advantaged some students while disadvantaging others. This is in line with previous literature that demonstrated common active learning practices, such as group work (Driessen, Knight, et al., 2020) increases student performance on average (Carmichael, 2009; Chaplin, 2009; Daniel, 2016; Donovan et al., 2018; Knight & Wood, 2005; Marbach-Ad et al., 2016; Springer et al., 1999; Weir et al., 2019; Yapici, 2016) but may disadvantage LGBTQIA+ students (Cooper & Brownell, 2016) and students with disabilities (Gin et al., 2020). Therefore, it is important for instructors to be deliberate about their teaching practices and aware of their students' needs when making pedagogical decisions.

Study Habits

Students who did and who did not face struggle frequently mentioned Study Habits in their explanation. Of those who struggled over the semester, students who overcame were often those who cited Study Habits as reason why they struggled. Students who could not overcome were less likely to cite Study Habits; rather, they characterized their struggle differently. Student responses overall often tied struggle to academic performance in the class. Given how important study habits are to performance outcomes (Numan & Hasan, 2017), the citation of Study Habits by both those who struggled or did not struggle makes sense, as more prepared students did better on exams, and exam scores were the bulk of students' grades. In introductory biology courses, previous work has shown the impact of incoming preparation on student performance (Salehi et al., 2020), and part of academic preparation relates to studying effectively for college-level exams. Future work will be needed to clarify the extent that a lesson on study habits and expectations have on students' experiences of struggle, academic performance and anxiety.

Prior Exposure to Biology

Students with prior biology courses and experiences often did not face struggle in our sample; however, previous literature concerning prior biology knowledge and student performance yields mixed messages. For example, Bone and Reid (2011) demonstrated prior experience with high school biology did not predict performance in a first-year biology course alone. That is, students who completed biology at the senior high school-level did perform better than those who had not, but only if they also completed chemistry (Bone & Reid, 2011). Similarly, Johnson and Lawson (1998) suggested reasoning ability contributes more to student performance in biology than prior biology knowledge. On the other hand, Loehr et al. (2012) claimed student performance in introductory college biology is positively associated with advanced high school science and mathematics coursework, an emphasis on a deep conceptual understanding of biology concepts, and a prior knowledge of concepts addressed in well-structured laboratory investigations. Similarly, Ozuru, Dempsey, and McNamara (2009), showed undergraduate students' prior biology knowledge was positively correlated with their overall comprehension of their college biology text.

Regardless of the mixed findings from previous literature, in this study it was clear that students often attributed their lack of struggle to prior preparation in biology, and if a student did not struggle, they were more likely to perform higher than predicted. This means that students' varying levels of exposure to biology could be contributing to a difference in ultimate performance, becoming more an effect of their social and academic capital from the time they were children than of an effect of college instruction (Marjoribanks, 1997). Previous research conducted by Salehi et al. (2019) demonstrated this to be true in the case of performance in standard introductory calculus-based mechanics. That is, incoming preparation predicted 20-30% of the variation in student exam performance. To circumvent this issue, we recommend instructors design their courses and teaching methods to better match the actual preparation level of their incoming students, rather than potentially assuming their students have previously learned the information in high school. This can be accomplished through a pre-course concept inventory, and the utilization of this information by the instructor may eliminate performance gaps while improving the success of all students.

Internal Versus External Attributions of Struggle or Lack of Struggle

Overall, students more commonly cited external sources (i.e., Prior STEM, Prior Biology, COVID-19, External Resources, and Classroom Factors) as the reason(s) they did or did not struggle rather than internal factors (i.e., Study Habits, Innate Ability, Time Management, Preference for Biology, and Anxiety). While the majority of student responses on the cause of struggle pointed to external factors, *more* students reported overcoming internal (rather than external) factors.

These results are largely consistent with previous literature on both mindset ("growth mindset" vs. "fixed mindset") and grit (perseverance and passion for long-term goals) (Duckworth et al., 2007; Limeri et al., 2020). The way students view their capacity for academic growth and change likely plays a large role in how they approach struggle, particularly in a STEM field where many students think of learning as an innate skill.

Grit theory predicts that a student's success is not only due to their capacity to learn but is largely predicted by measures of a student's determination. We acknowledge that this theory has serious limitations in its application, and we do not intend to shift away responsibility from situational factors that legitimately challenge students (Ris, 2015). Additionally, we acknowledge that this study is conducted at a school that disproportionately serves middle to upper-class students, and is among the lowest in enrollment of students from the bottom 20%

(Aisch et al., 2017). Instead, we recognize that introductory courses are required to progress within the major for science students. Because we observed that students are more likely to overcome internal struggle, we suggest that interventions targeting mindset and/or grit may improve students' ability to overcome challenges they face (Binning et al., 2019, 2020; Harackiewicz & Priniski, 2018; Limeri et al., 2020).

Additionally, these divisions among students based on whether they experienced challenges of introductory courses and whether they could overcome them reflects well-documented research on persistence and loss in undergraduate STEM education. For example, Seymour and Hunter (2019) described different types of processes that accounted for student decisions to leave STEM majors. Tracking students over time revealed a "push-pull" decision-making process, where students simultaneously experienced "push" factors such as problems in students' precollege and college experiences, as well as "pull" factors, or perceived attractions of alternative majors or career trajectories. First, they showed that students' decision to switch majors were due primarily to external factors such as problems in course design, teaching, and negative classroom culture, these are the same external factors to which our students largely attributed their sources of struggle. Second, Seymour and Hunter (2019) observed that students who persisted in a major developed coping strategies to help them persevere. Since our study does not investigate the coping strategies used by students who overcame their struggle, we acknowledge that exploring coping strategies in biology among students who overcame struggle is the next logical step of the current research.

LIMITATIONS

This study has several limitations. First, we assumed students interpreted the word 'struggle' in the same way. An advantage of leaving the term open for interpretation was that students communicated in their responses how they interpreted struggle and how they viewed struggle in the classroom. Based on student responses, struggle may have been interpreted as challenges impacting performance in the course or the ability to learn.

As this study took place over two consecutive semesters with some courses which are often taken consecutively (Introductory Biology I in Fall 2019 and Introductory Biology II in Spring 2020), there were 130 students with duplicate survey entries (260 entries out of 965) and thus our dataset does not have completely independent samples between semesters. While this does not represent completely independent sampling, we chose to retain all duplicate entries for our study, and control for duplicate student responses in the quantitative portion by including student ID as a random effect.

Lastly, while struggle can significantly undermine students' academic abilities and performance (Batz et al., 2015; England et al., 2017b), struggle disproportionately affects PEERs, who already face unique challenges resulting from social isolation and discrimination (Bertrand & Mullainathan, 2004; Hurtado et al., 2010). We acknowledge that the students who participated in our study are largely from over-represented backgrounds (Table S2), necessitating a repeat study with students from low socioeconomic backgrounds or with PEERs to target those students in order to make sure introductory biology courses do not effectively maintain and contribute to social and cultural inequities. Ultimately, promoting student retention in STEM is essential to our national efforts to produce graduates that meet the growing need for a trained and diverse workforce (President's Council of Advisors on Science and Technology, 2012), a goal which may be addressed through mitigation of student struggle in STEM courses.

CONCLUSIONS

Students reported struggle across introductory biology for a variety of reasons, as demonstrated in our findings. Quantitative and qualitative analyses revealed that students' interpretation of struggle was interwoven with their perceptions of performance or learning in the class. We found that obstacles experienced by some students served as life rafts for other students. For this reason, it is important for instructors and researchers alike to know about the common pitfalls some introductory biology students face, so they can intervene with recommended strategies, creating more equitable classrooms where all students can succeed, regardless of prior experience with biology, knowledge of effective study habits, and classroom elements.

ACKNOWLEDGEMENTS

The qualitative survey used for this study was adapted from a survey developed by Dr. Lisa Limeri, who also provided interesting discussion on the project in an Introduction to Discipline-Based Education Research course taught by CJB. We would also like to thank Dr. Todd Steury for his wisdom on quantitative analyses. Finally, we would like to thank all of the members of the Ballen lab for their feedback on the paper. Graphs were created in R and many figures were formatted with BioRender.com. This work was supported by NSF DUE-2120934 awarded to AEB and CJB; DUE-2011995 and DBI-1919462 awarded to CJB.

References

Aisch, G., Buchanan, L., Cox, A., & Quealy, K. (2017). Some colleges have more students from the top 1 percent than the bottom 60. Find yours. *The New York Times*.

Ali, I., Narayan, A. K., & Sharma, U. (2020). Adapting to COVID-19 disruptions: student engagement in online learning of accounting. *Accounting Research Journal*. https://doi.org/10.1108/ARJ-09-2020-0293

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 Sciences Education*, *8*(3), 203–213. https://doi.org/10.1187/cbe.09-03-0025

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. *Journal of Experimental Social Psychology*, *38*(2), 113–125. https://doi.org/10.1006/JESP.2001.1491

Austin, A. C., Hammond, N. B., Barrows, N., Gould, D. L., & Gould, I. R. (2018). Relating motivation and student outcomes in general organic chemistry. *Chemistry Education Research and Practice*, *19*(1), 331–341. https://doi.org/10.1039/C7RP00182G

Ballen, C. J., Salehi, S., & Cotner, S. (2017). Exams disadvantage women in introductory biology. *PLoS ONE*, *12*(10), e0186419. https://doi.org/10.1371/journal.pone.0186419

Batz, Z., Olsen, B. J., Dumont, J., Dastoor, F., & Smith, M. K. (2015). Helping struggling students in introductory biology: A peer-tutoring approach that improves performance, perception, and retention. *CBE Life Sciences Education*, *14*(2), ar16. https://doi.org/10.1187/cbe.14-08-0120

Beatty, A. E., Esco, A., Curtiss, A. B. C., & Ballen, C. J. (2022). Students who prefer face-to-face tests outperform their online peers in organic chemistry. *Chemistry Education Research and Practice*. https://doi.org/10.1039/D1RP00324K

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991–1013. https://doi.org/10.1257/0002828042002561

Binning, K. R., Kaufmann, N., McGreevy, E. M., Fotuhi, O., Chen, S., Marshman, E., Kalender, Z. Y., Limeri, L. B., Betancur, L., & Singh, C. (2020). Changing Social Contexts to Foster Equity in College Science Courses: An Ecological-Belonging Intervention. *Psychological Science*, *31*(9), 1059–1070. https://doi.org/10.1177/0956797620929984

Binning, K. R., Wang, M. Te, & Amemiya, J. (2019). Persistence Mindset among Adolescents: Who Benefits from the Message that Academic Struggles are Normal and Temporary? *Journal of Youth and Adolescence*, *48*(2), 269–286. https://doi.org/10.1007/s10964-018-0933-3

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

Bone, E. K., & Reid, R. J. (2011). Prior learning in biology at high school does not predict performance in the first year at university. *Higher Education Research and Development*, *30*(6), 709–724. https://doi.org/10.1080/07294360.2010.539599

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances*, *5*(2). https://doi.org/10.1126/SCIADV.AAU4734

Canning, E. A., Ozier, E., Williams, H. E., AlRasheed, R., & Murphy, M. C. (2021). Professors Who Signal a Fixed Mindset About Ability Undermine Women's Performance in STEM. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506211030398

Carmichael, J. (2009). Team-based learning enhances performance in introductory biology.

Journal of College Science Teaching, *38*, 54–61.
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). https://doi.org/10.1187/cbe.17-06-0107
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*, 72–79.
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). https://doi.org/10.1187/cbe.16-01-0074
Cooper, K. M., & Brownell, S. E. (2020). Student Anxiety and Fear of Negative Evaluation in Active Learning Science Classrooms. In *Active Learning in College Science* (pp. 909–925). https://doi.org/10.1007/978-3-030-33600-4_56
Corkin, D. M., Horn, C., & Pattison, D. (2017). The effects of an active learning intervention in biology on college students' classroom motivational climate perceptions, motivation, and achievement. *Educational Psychology*, *37*(9), 1106–1124. https://doi.org/10.1080/01443410.2017.1324128
Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable? *PLOS ONE*, *12*(12), e0189610. https://doi.org/10.1371/JOURNAL.PONE.0189610
Cromley, J. G., Perez, T., & Kaplan, A. (2016). Undergraduate STEM Achievement and Retention. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 4–11. https://doi.org/10.1177/2372732215622648
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
DeFeo, D. J., Tran, T. C., & Gerken, S. (2021). Mediating Students' Fixation with Grades in an Inquiry-Based Undergraduate Biology Course. *Science and Education*, *30*(1), 81–102. https://doi.org/10.1007/s11191-020-00161-3
Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257. https://doi.org/10.1073/PNAS.1821936116
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). https://doi.org/10.1187/cbe.17-12-0283
Driessen, E., Beatty, A., Stokes, A., Wood, S., & Ballen, C. (2020). Learning principles of evolution during a crisis: An exploratory analysis of student barriers one week and one month into the COVID-19 pandemic. *Ecology and Evolution*, *10*(22), 12431–12436. https://doi.org/10.1002/ECE3.6741
Driessen, E., 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. https://doi.org/10.1187/cbe.20-04-0068
Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087
Dweck, C. S. (1999). *Self-theories : their role in motivation, personality, and development*.

Dyrberg, N. R., & Holmegaard, H. T. (2019). Motivational patterns in STEM education: a self-determination perspective on first year courses. *Research in Science and Technological Education*, *37*(1), 90–109. https://doi.org/10.1080/02635143.2017.1421529

England, B. J., Brigati, J. R., & Schussler, E. E. (2017a). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PLOS ONE*, *12*(8), e0182506. https://doi.org/10.1371/JOURNAL.PONE.0182506

England, B. J., Brigati, J. R., & Schussler, E. E. (2017b). Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major. *PLOS ONE*, *12*(8), e0182506. https://doi.org/10.1371/journal.pone.0182506

Flanagan, K. M., & Einarson, J. (2017). Gender, math confidence, and grit: Relationships with quantitative skills and performance in an undergraduate biology course. *CBE Life Sciences Education*, *16*(3). https://doi.org/10.1187/cbe.16-08-0253

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*, *111*(23), 8410–8415. https://doi.org/10.1073/PNAS.1319030111

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., & Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. *CBE Life Sciences Education*, *6*(2), 132–139. https://doi.org/10.1187/cbe.06-09-0194

Gainen, J. (1995). Barriers to success in quantitative gatekeeper courses. *New Directions for Teaching and Learning*, *61*, 5–14.

Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., & Chang, M. J. (2012). From Gatekeeping to Engagement: A Multicontextual, Mixed Method Study of Student Academic Engagement in Introductory STEM Courses. *Research in Higher Education*, *53*(2), 229–261. https://doi.org/10.1007/s11162-011-9247-y

Gin, L. E., Guerrero, F. A., Cooper, K. M., & Brownell, S. E. (2020). Is active learning accessible? Exploring the process of providing accommodations to students with disabilities. *CBE Life Sciences Education*, *19*(4), 1–15. https://doi.org/10.1187/cbe.20-03-0049

Glesne, C., & Peshkin, A. (1992). *Becoming Qualitative Researchers: An Introduction*.

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving Student Outcomes in Higher Education: The Science of Targeted Intervention. In *Annual Review of Psychology* (Vol. 69, pp. 409–435). Annual Reviews Inc. https://doi.org/10.1146/annurev-psych-122216-011725

Hurtado, S., Newman, C. B., Tran, M. C., & Chang, M. J. (2010). Improving the rate of success for underrepresented racial minorities in STEM fields: Insights from a national project. *New Directions for Institutional Research*, *2010*(148), 5–15. https://doi.org/10.1002/IR.357

Johnson, D. R., Soldner, M., Leonard, J. B., Alvarez, P., Inkelas, K. K., Rowan-Kenyon, H. T., & Longerbeam, S. D. (2007). Examining sense of belonging among first-year undergraduates from different racial/ethnic groups. *Journal of College Student Development*, *48*(5), 525–542. https://doi.org/10.1353/csd.2007.0054

Johnson, M. A., & Lawson, A. E. (1998). What Are the Relative Effects of Reasoning Ability and Prior Knowledge on Biology Achievement in Expository and Inquiry Classes? *J Res Sci Teach*, *35*, 89–103. https://doi.org/10.1002/(SICI)1098-2736(199801)35:1

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(WINTER), 298–310. https://doi.org/10.1187/05-06-0082

Lewis, C. S., Williams, B. Della, Sohn, K., & Loy, T. C. (2017). The Myth of Entitlement: Students' Perceptions of the Relationship between Grading and Learning at an Elite University. *The Qualitative Report*, *22*(11), 2997–3010.

Limeri, L. B., Carter, N. T., Choe, J., Harper, H. G., Martin, H. R., Benton, A., & Dolan, E. L. (2020). Growing a growth mindset: characterizing how and why undergraduate students' mindsets change. *International Journal of STEM Education*, *7*(1), 1–19. https://doi.org/10.1186/s40594-020-00227-2

Loehr, J. F., Almarode, J. T., Tai, R. H., & Sadler, P. M. (2012). High school and college biology: A multi-level model of the effects of high school courses on introductory course performance. *Journal of Biological Education*, *46*(3), 165–172. https://doi.org/10.1080/00219266.2011.617767

Marbach-Ad, G., Rietschel, C. H., Saluja, N., Carleton, K. L., & Haag, E. S. (2016). The Use of Group Activities in Introductory Biology Supports Learning Gains and Uniquely Benefits High-Achieving Students. *Journal of Microbiology & Biology Education*, *17*(3), 360–369. https://doi.org/10.1128/jmbe.v17i3.1071

Marjoribanks, K. (1997). Family Background, Social and Academic Capital, and Adolescents' Aspirations: A Mediational Analysis. *Social Psychology of Education 1997 2:2*, *2*(2), 177–197. https://doi.org/10.1023/A:1009602307141

Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm? *Physical Review Physics Education Research*, *14*(2), 020123. https://doi.org/10.1103/PhysRevPhysEducRes.14.020123

Meaders, C. L., Lane, A. K., Morozov, A. I., Shuman, J. K., Toth, E. S., Stains, M., Stetzer, M. R., Vinson, E., Couch, B. A., & Smith, M. K. (2020). Undergraduate Student Concerns in Introductory STEM Courses: What They Are, How They Change, and What Influences Them. *Journal for STEM Education Research*, *3*(2), 195–216. https://doi.org/10.1007/s41979-020-00031-1

Mervis, J. (2011). Weed-Out Courses Hamper Diversity. *Science*, *334*(6061), 1333–1333. https://doi.org/10.1126/SCIENCE.334.6061.1333

Morgan, D. L. (1993). Qualitative content analysis: a guide to paths not taken. *Qualitative Health Research*, *3*(1), 112–121. https://doi.org/10.1177/104973239300300107

Numan, A., & Hasan, S. S. (2017). Effect of Study Habits on Test Anxiety and Academic Achievement of Undergraduate Students. *Journal of Research and Reflections in Education*, *11*(1), 1–14.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228–242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Ris, E. W. (2015). Grit: A Short History of a Useful Concept. *Journal of Educational Controversy*, *10*(1), 1–18.

Ryan, R. M., & Deci, E. L. (2017). Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. In *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press. https://doi.org/10.1521/978.14625/28806

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sabot, R., & Wakeman-Linn, J. (1991). Grade Inflation and Course Choice. *Journal of Economic Perspectives*, *5*(1), 159–170. https://doi.org/10.1257/jep.5.1.159

Saldaña, J. (2009). The Coding Manual for Qualitative Researchers. In *SAGE Publications Inc.: Vol. XXXIII* (Issue 2).

Salehi, S., Burkholder, E., Lepage, G. P., Pollock, S., & Wieman, C. (2019). Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics. *Physical Review Physics Education Research*, *15*(2). https://doi.org/10.1103/PhysRevPhysEducRes.15.020114

Salehi, S., Cotner, S., & Ballen, C. J. (2020). Variation in Incoming Academic Preparation: Consequences for Minority and First-Generation Students. *Frontiers in Education*, *5*, 552364. https://doi.org/10.3389/FEDUC.2020.552364

Sambell, K., & Hubbard, A. (2004). The Role of Formative 'Low-stakes' Assessment in Supporting Non-traditional Students' Retention and Progression in Higher Education: Student Perspectives. *Widening Participation and Lifelong Learning*, *6*(2), 25–36.

Scott, A., McNair, D., Lucas, J., & Land, K. (2017). Research and Teaching: From Gatekeeper to Gateway: Improving Student Success in an Introductory Biology Course. *Journal of College Science Teaching*, *046*(04). https://doi.org/10.2505/4/jcst17_046_04_93

Seidel, S. B., Reggi, A. L., Schinske, J. N., Burrus, L. W., & Tanner, K. D. (2015). Beyond the biology: A systematic investigation of noncontent instructor talk in an introductory biology course. *CBE Life Sciences Education*, *14*(4). https://doi.org/10.1187/cbe.15-03-0049

Seymour, E., & Hunter, A.-B. (2019). *Talking about leaving revisited : persistence, relocation, and loss in undergraduate STEM education*.

Sheshadri, A., Gitinabard, N., Lynch, C. F., Barnes, T., & Heckman, S. (2019). Predicting student performance based on online study habits: A study of blended courses. In *arXiv*.

Sithole, A., Chiyaka, E. T., McCarthy, P., Mupinga, D. M., Bucklein, B. K., & Kibirige, J. (2017). Student Attraction, Persistence and Retention in STEM Programs: Successes and Continuing Challenges. *Higher Education Studies*, *7*(1), 46. https://doi.org/10.5539/hes.v7n1p46

Smoyer, A. B., O'Brien, K., & Rodriguez-Keyes, E. (2020). Lessons learned from COVID-19: Being known in online social work classrooms. *International Social Work*, *63*(5), 651–654. https://doi.org/10.1177/0020872820940021

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

Supriya, K., Mead, C., Anbar, A. D., Caulkins, J. L., Collins, J. P., Cooper, K. M., LePore, P. C., Lewis, T., Pate, A., Scott, R. A., & Brownell, E. (2021). COVID-19 and the abrupt shift to remote learning: Impact on grades and perceived learning for undergraduate biology students. *BioRxiv*, 5–6. https://doi.org/10.1101/2021.03.29.437480

Theobald, E. (2018). Students Are Rarely Independent: When, Why, and How to Use Random Effects in Discipline-Based Education Research. *CBE—Life Sciences Education*, *17*(3), rm2. https://doi.org/10.1187/cbe.17-12-0280

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., 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

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), e0220900. https://doi.org/10.1371/JOURNAL.PONE.0220900

Wester, E. R., Walsh, L. L., Arango-Caro, S., & Callis-Duehl, K. L. (2021). Student Engagement Declines in STEM Undergraduates during COVID-19–Driven Remote Learning. *Journal of Microbiology & Biology Education*, *22*(1). https://doi.org/10.1128/JMBE.V22I1.2385

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. https://doi.org/10.1187/cbe.18-08-0155

Wolinsky, H. (2021). Biology lessons in times of COVID. *EMBO Reports*, *22*(3), e52513. https://doi.org/10.15252/EMBR.202152513

Wut, T., & Xu, J. (2021). Person-to-person interactions in online classroom settings under the impact of COVID-19: a social presence theory perspective. *Asia Pacific Education Review 2021*, *1*, 1–13. https://doi.org/10.1007/S12564-021-09673-1

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). https://doi.org/10.11114/jets.v4i4.1372

Ye, L., Shuniak, C., Oueini, R., Robert, J., & Lewis, S. (2016). Can they succeed? Exploring at-risk students' study habits in college general chemistry. *Chemistry Education Research and Practice*, *17*(4), 878–892. https://doi.org/10.1039/c6rp00101g

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature 2019 573:7774*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

Table 1: Participating classroom details

Class	***Instructor***	***Semester***	***Course Subject***	***Participating students****	***Total students***
1	*1*	*Fall 2019*	*Course 1*	*205*	*327*
2	*2*	*Fall 2019*	*Course 1*	*107*	*159*
3	*2*	*Fall 2019*	*Course 1*	*135*	*260*
4	*3 & 4*	*Spring 2020*	*Course 2*	*172*	*237*
5	*3 & 4*	*Spring 2020*	*Course 2*	*137*	*194*
6	*5*	*Spring 2020*	*Course 3*	*209*	*376*

Variety in student response rates across the six classes depend on student burnout at the end of the semester (i.e., when our data was collected) as well as instructor enthusiasm towards encouraging their students to take the survey.

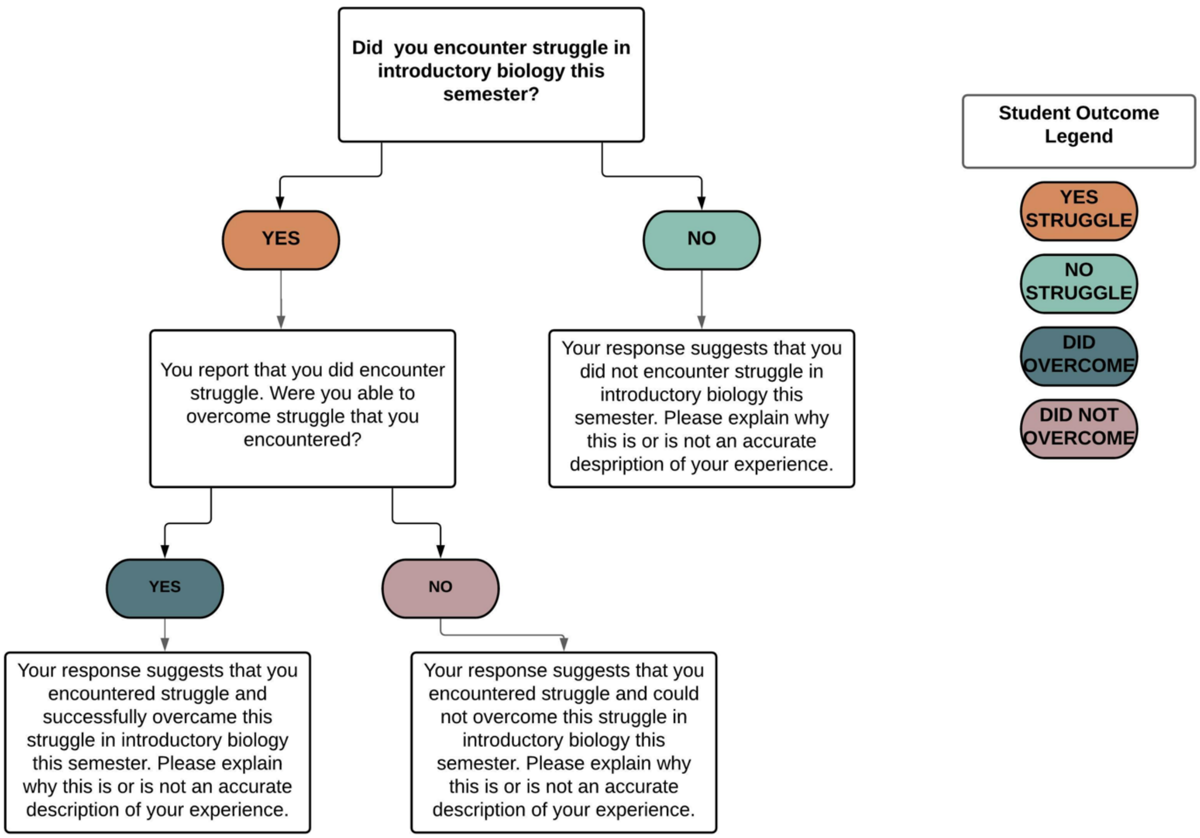

Figure 1. Logic question progression based upon student responses to the question "Did you encounter struggle in introductory biology this semester?" The resulting responses led to the following student outcomes: (1) No, I did not struggle and (2) Yes, I did struggle. If the student indicated Yes, then they were further asked "Were you able to overcome the struggle that you encountered?". The resulting responses led to the following outcomes: (3) No, I did not overcome my struggle and (4) Yes, I did overcome my struggle.

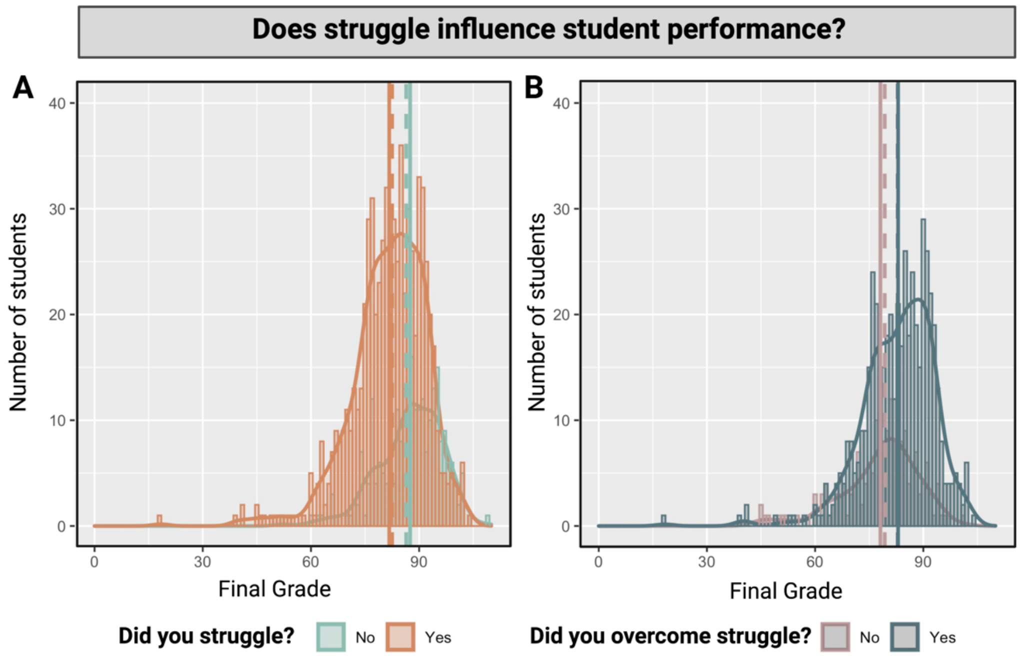

Figure 2. Histogram of final scores grouped by (A) whether students did or did not encounter struggle, and (B) whether students who encountered struggle did or did not overcome their struggle. Hashed line represents the estimated marginal means in final score for each group, and solid lines represent raw means in final score for each group. The estimated marginal mean is the mean of the final score for each group of struggle (y/n) at the mean value of incoming preparation (PC1) based on the model.

Categories	Codes	Code Explanation	Student Excerpt Example: Positive	Student Excerpt Example: Negative
External: Outside of a student's control (student is not the subject)	Prior STEM	The student mentions prior experience or lack of experience with non-biology STEM-related material.	"I did not struggle with the class as I am a Junior in Chemical Engineering.	N/A
	Prior Biology	The student mentions prior experience or lack of experience with biology-related material.	"I have taken countless biology courses."	N/A
	COVID-19	The student indicates that COVID-19 caused difficulty. Student's must make a direct reference to "COVID19" or the "pandemic" or "quarantine."	"I had a real problem memorizing the different phylogenetic trees, but all the extra time I had as a result of the quarantine allowed we to really practice and nail them down."	"I encountered struggle while dealing with the coronavirus situation. It was just a difficult and confusing time for me, but as my family member started getting better, it was somewhat easier to focus."
	External Resources	Resources that are provided by the instructor/institution but are used outside of class time. This includes attending office hours, SI sessions, or getting help from learning assistants and/or teaching assistants.	"The TAs were very helpful when I had a question."	"This course was very hard and I was not taught the material very well. I wish my LA had helped me more."
	Classroom Factors	References the general classroom environment. See classroom reference subcode for further breakdown.	"The information was fun and interesting to learn and the teachers provided multiple opportunities to retain the information, so this allowed me to avoid struggle."	"I struggled with assignments that were not announced to be completed and the failure to announce what to review before classes."
Internal: Within a student's control	Study Habits	Student directly mentions a negative or positive effect of study habits. This code includes terminology such as "preparing for exams" and/or "struggling with memorization". This also includes any study tactics used such as YouTube videos, study groups, etc.	"While those around me struggled, I found that merely reading the slides and doing the guided reading coupled with the Mastering Biology assignments and quizzes was enough for me."	"I struggled because I was not able to find the proper way to study."
	Innate Ability	The student talks about their innate ability to perform as a biologist, a student, or a thinker.	"I enjoy biology and am naturally good at it."	"Biology is a subject that takes longer for me to comprehend than any other subject."
	Time Management	Student relates their reported struggles to an inability or ability to manage their obligations within their time availability. This can include mentions issues in their personal life.	N/A	"I had to manage time between this class principals of chemical engineering and calculus 3."
	Preference for Biology	Student mentions preference for and/or enjoyment of biology. Alternatively, student mentions they do not prefer or enjoy the subject.	"I really enjoy biology."	"I suck at biology and have 0 interest in it so why am I forced to take a class on it when it has nothing to do with what I want to do in my career?!"
	Anxiety	Student mentions the presence or a lack of the presence of anxiety, stress, nerves, or depression associated with the course.	"I never felt overwhelmed. The class is generally what I expected."	"I felt unprepared and overwhelmed at the beginning."

Figure 3. Explanations of ten qualitative codes with examples of a positive and negative student response for each code.

Codes	Code Explanation	Student Excerpt Example	
		Positive	Negative
Instructor	Mentions the instructor in a way that is directly affecting their learning experience.	"I felt my professor prepared us and I worked very hard outside of class."	"The professor just didn't click with me."
Exams	Mentions the exam as either their reason for struggle or lack of struggle.	"I felt that I was very prepared for tests/quizzes and I understood the material being taught to me."	"The struggles were with that the tests were very hard but I had to adjust my studying to perform better."
Format	Mentions classroom style, flipped classroom, or class size as the reason for their struggle or lack of struggle.	"The setup of the class allowed me to encounter the same information multiple times in different ways and thus easily retain it."	"It was very difficult for me to learn in a classroom that was flipped."
Content	Mentions material or content being the reason they struggled or did not struggle. This could include specific material or content references, including phylogenetics or evolution, etc.	"I understood most concepts covered in the class."	"The course material was very complex and difficult, yet we had to learn it all so quickly."
Group Work	Mentions group work in the class as the source of challenge or struggle. This includes mentioning peer learning.	"I had really good group members so if I was confused about anything, they helped it not turn into a struggle."	"The primary struggle I had was the group work specific to this class rather than the actual academic content."
Workload	The student mentions class workload, including abundancy of materials and pace of delivery in reference to an external difficulty seen as out of their direct control.	N/A	"I had a hard time with the fast pace of the class."

Figure 4. Breakdown of six "Classroom Factors" subcodes with examples of a positive and negative student response for each.

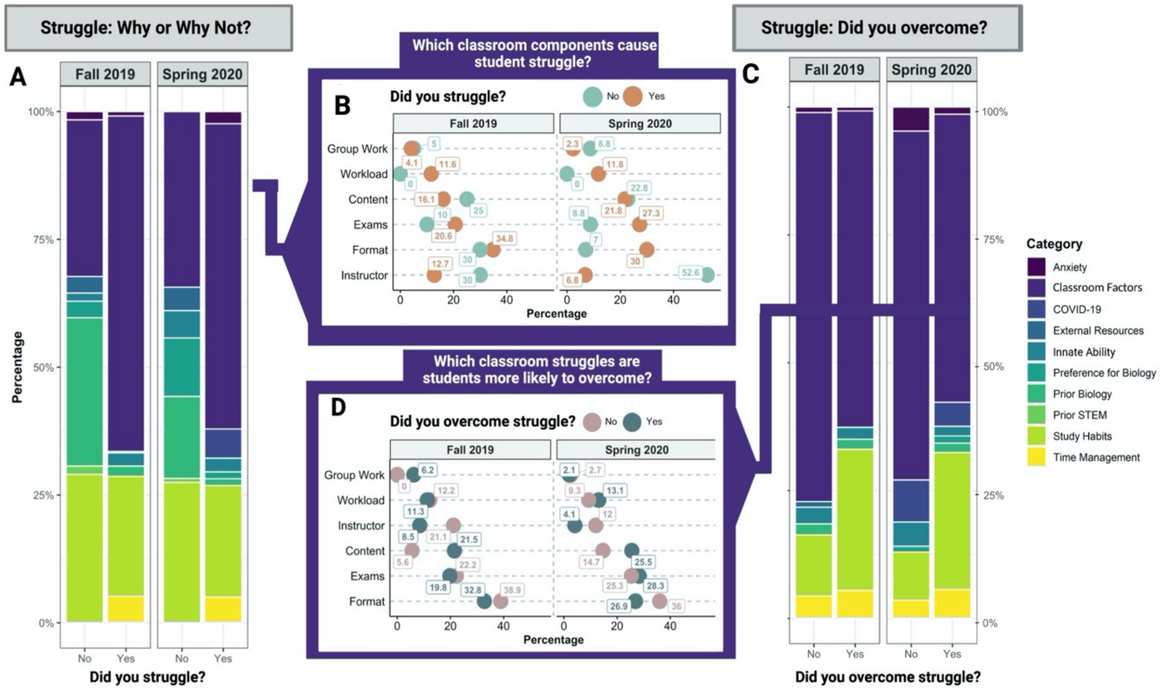

Figure 5: Fall 2019 and Spring 2020 student responses as categorized into the ten codes. **A**) Student responses indicate the source of struggle (Yes), or reason for a lack of struggle (No) across semesters. **B**) Of students who reported struggle, responses reflect source of struggle among those who overcame (Yes) and could not overcome (No) across semesters. **C**) Breakdown of category "Classroom Factors" as a source of struggle or lack of struggle into six subcodes displayed in ascending order. **D**) Of students who reported struggle, breakdown of category "Classroom Factors" as a source of struggle among students who could overcome and those who could not into six subcodes displayed in ascending order.

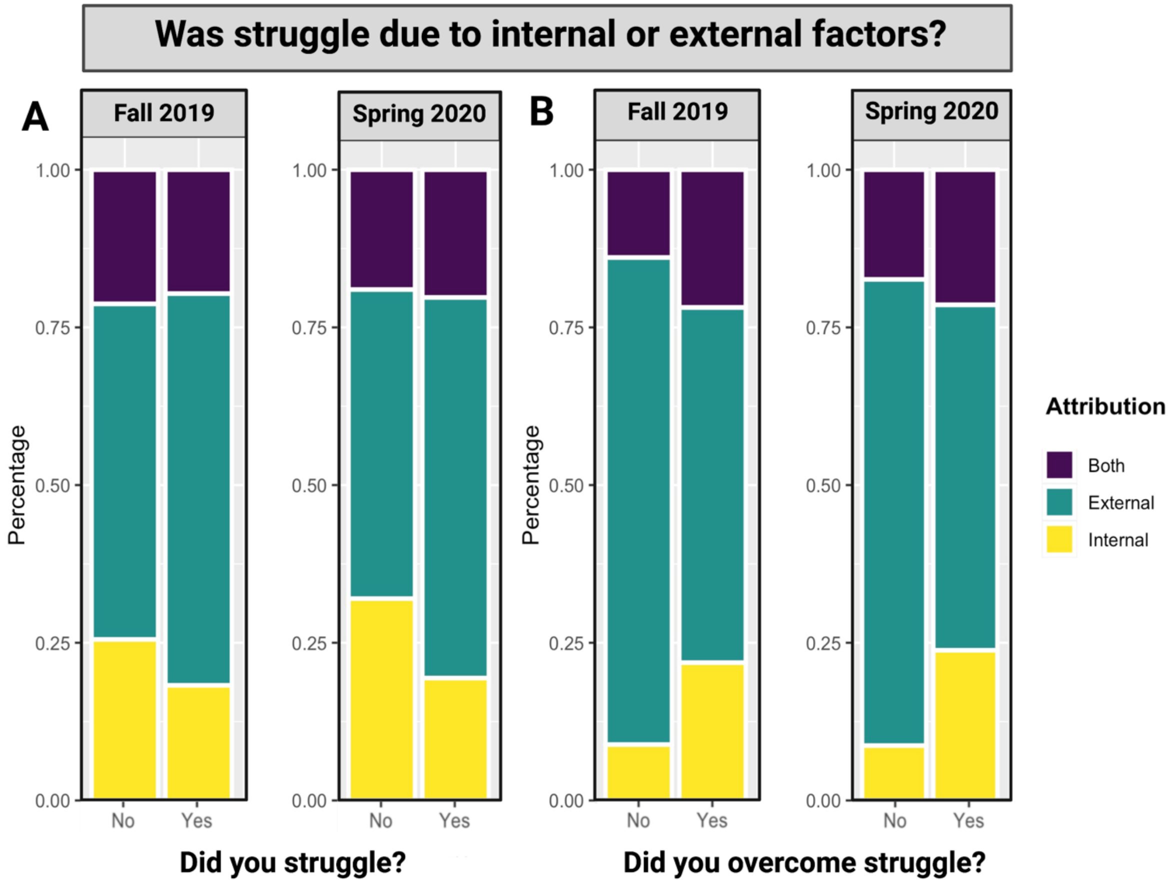

Figure 6. Fall 2019 and Spring 2020 student responses as collapsed into the two overarching categories: externally or internally attributed factors leading either to (A) students reporting they did not experience struggle (No) or that they experienced struggle (Yes); (B) Of students who experienced struggle, students reporting they were not able to overcome struggle (No) or that they could overcome struggle (Yes). Of note, external factors include Prior STEM (outside of biology), Prior Biology, COVID-19, External Resources, and Classroom Factors codes and internal factors represent Study Habits, Innate Ability, Time Management, Preference for Biology, and Anxiety. External factors are those outside of the students' control and
Internal factors are those within the students' control.