A Course-Based Undergraduate Research Experience Improves Outcomes in Mentored Research

Justin Fendos,[†#]* Liang Cai,[†] Xianmei Yang,[†] Guodong Ren,[†] Lin Li,[†] Zhiqiang Yan,[†§] Boxun Lu,[†] Yan Pi,[†] Jinbiao Ma,[†] Bin Guo,[†] Xiaohui Wu,[†] Pingli Lu,[†‖] Ruilin Zhang,[†#] and Ji Yang[†]

[†]School of Life Sciences, Fudan University, Shanghai 200438, China; [‡]School of Energy and Life Sciences, Dongseo University 47017, South Korea; [§]Institute of Molecular Physiology, Shenzen Bay Laboratory, Shenzhen 510852, China; [‖]School of Life Sciences, Henan University, Kaifeng 475001, China; [#]School of Basic Medical Sciences, Wuhan University, Wuhan 430072, China

ABSTRACT

Infusing undergraduate curricula with authentic research training is an important contemporary challenge. Such exposure typically occurs through mentored research (MR) or course-based undergraduate research experiences (CUREs). In Asian contexts, CURE implementation is rare, while MR is often a graduation requirement. In this study, mentor interviews and mentee focus groups were used to characterize the learning challenges associated with this requirement at a Chinese university. An intensive 6-week CURE was then implemented as an MR preparatory program to help mitigate the identified challenges. This program contained seven site-specific features not typically included in other CUREs, each designed to improve different aspects of student readiness for MR. Post-CURE surveys, focus groups, and interviews demonstrated CURE enrollment significantly improved subsequent MR outcomes. Almost 90% of all enrollees, for example, began their first MR experience in their second year, more than twice the rate of non-enrollees. Enrollees also reported greater confidence in their research skills and more frequent experiences working in multiple labs. This study reports both immediate CURE and downstream MR outcomes, using the former to help explain the latter. A comprehensive CURE implementation process is described, offering a potential model for the design of other programs with similar research enhancement goals.

INTRODUCTION

The need to facilitate research training in higher education has been a consistent point of emphasis in recent years, especially in the United States (National Research Council, 2003; President's Council of Advisors on Science and Technology, 2012). The challenge of integrating this training into undergraduate curricula is a difficult one, necessitating the development of many complex skills rather than mere transfers of conceptual knowledge (Baker *et al.*, 2015; Brownell *et al.*, 2015). Across the globe, mentored research (MR) is often the primary mode of research skill acquisition, a practice that requires undergraduates to independently seek their own opportunities by joining a university-affiliated lab (Hunter *et al.*, 2007; Russell *et al.*, 2007; American Association for the Advancement of Science, 2011). In Asian contexts, these experiences are usually termed "apprenticeship-based" (Thiry and Laursen, 2011; Wei and Woodin, 2011; Hung *et al.*, 2015).

In addition to providing important learning benefits, MR offers the opportunity for faculty to harness undergraduates to increase the volume of publishable work (Hunter *et al.*, 2007; Dolan and Johnson, 2010; Linn *et al.*, 2015). In virtually every country, science, technology, engineering, and mathematics (STEM) labs often lack professional researchers, resulting in a high reliance on students—both

graduate and undergraduate—for productivity. In many Asian contexts, graduate students are either absent or systematically limited, precipitating a corresponding dependence on undergraduates.

To encourage MR, many Asian universities have implemented, either independently or via government recommendation, graduation requirements necessitating undergraduate participation, often for a minimum of a semester or academic year (Choi *et al.*, 2011; Wei and Woodin, 2011; Abu-Zaid and Alkattan, 2013). Some institutions also have requirements for the volume of work that must be completed, imposing specific demands on productivity. Because STEM departments typically have many more undergraduates than research groups, MR recruitment is often unregulated, giving faculty the theoretical opportunity to employ an unlimited number of apprentices. In this manner, MR is often an integral and competitive aspect of productivity at many Asian universities, especially critical for those with few or no graduate students.

In recent years, the United States and a host of other countries have prioritized STEM education on a national level, sparking great interest in the development of effective curricular designs for research training (Spell *et al.*, 2014; Linn *et al.*, 2015; Ballen *et al.*, 2017; Jones and Lerner, 2019; McGill *et al.*, 2019). In the life sciences, these methods have included the integration of authentic research activities into core courses (Auchincloss *et al.*, 2014; Brownell and Kloser, 2015; Linn *et al.*, 2015). These course-based undergraduate research experiences (CUREs) are defined by five characteristics (Auchincloss *et al.*, 2014; Brownell *et al.*, 2015; Brownell and Kloser, 2015): 1) elements of discovery that allow students to generate or work with novel data; 2) research activity iterations that help students build competence and confidence in targeted skills through repetition; 3) student collaboration when making research decisions or data interpretations; 4) an emphasis on implementing scientific methods and critical thinking; and 5) use of broadly relevant learning topics, so the resulting work may be publishable or of interest to others outside the learning environment. These characteristics, in essence, make CUREs *authentic*, mimicking the structure of professional research, an antithesis to "cookbook" lab courses, which are defined by a lack of discovery, iterations, and student ownership over the learning process (Cuthbert *et al.*, 2012; Brownell and Kloser, 2015; Rowland *et al.*, 2016).

Although a sizable body of work has investigated the learning challenges associated with MR in Western contexts (Dolan and Johnson, 2009, 2010; Wei and Woodin, 2011; Aikens *et al.*, 2016), no work has yet conducted a similar investigation in an Asian setting or one with MR graduation requirements. To this end, the present study employed mentor interviews and undergraduate focus groups to characterize these challenges at a Chinese university. This information was used to design an intensive 6-week CURE called BIOS, which was implemented as an MR preparatory program over a 5-year period. A series of post-program surveys, focus groups, and interviews demonstrated BIOS enrollment significantly improved subsequent participation, confidence, and performance in MR. As far as the authors are aware, this the first time a CURE has been shown to systematically improve MR outcomes across an entire population of life science undergraduates.

In summary, the present study investigated two research questions:

1. What challenges do stakeholders face when engaging in compulsory MR?
2. How can a CURE be used as a preparatory program to alleviate these concerns?

METHODS

Pre-CURE Focus Groups, Interviews

To characterize the challenges associated with compulsory MR, 25 undergraduates were recruited into three focus groups, while eight faculty and 12 postgraduates were individually interviewed (Supplemental Table 1). All participants were affiliates of the same School of Life Sciences at a highly competitive research university in Shanghai. All interactions were conducted in early 2015. Most faculty interviewees had fewer than 10 years of experience being a professor, while all postgraduates were either graduate students or professional research staff (Supplemental Table 1). Most undergraduates were third- or fourth-year students with prior or ongoing MR participation (Supplemental Table 1).

CURE Design Process

In recent years, several theoretical frameworks have been published to guide CURE design. Work by Bakshi and colleagues (2016), for example, offered activity recommendations for developing specific skills, while Brownell and Kloser (2015) offered strategies for assessing CURE effectiveness. Because neither had been published at the time of BIOS's inception in early 2015, BIOS was not designed by referencing them. Instead, its framework was constructed independently using two sets of principles: the five defining characteristics of a CURE, as described in the *Introduction*, and the MR-associated challenges revealed through pre-CURE interviews and focus groups.

Despite the alternative design process, BIOS's architecture nevertheless exhibited many similarities with the recommendations and strategies put forth by Bakshi *et al.* (2016) and Brownell and Kloser (2015). Like Bakshi *et al.*, BIOS made use of linear sequences of research tasks (e.g., hypothesis to experiment to data interpretation), allowing students to experience common steps in the scientific process. Other similarities included the use of writing activities before and after experiments (i.e., proposals and reports) and the frequent implementation of cooperative learning (Huitt *et al.*, 2015; Connell *et al.*, 2016), requiring students to work in pairs or groups to enhance interaction and feedback (Gillies, 2003; Stoltzfus and Libarkin, 2016; Warfa, 2016). Similarities to Brownell and Kloser (2015) included the use of direct observational assessments and pre–post surveys to determine changes in lab skill proficiencies and learning attitudes. Deviations included BIOS's lack of emphasis on primary literature and poster design. The former was avoided because the English abilities of Chinese undergraduates were very uneven, while the latter was avoided to allocate more time for experiments.

As described in Figure 1, BIOS contained six topical tracks: biochemistry, cell biology, fish genetics, fly genetics, mouse genetics, and plant physiology. Each had its own schedule of unique learning activities and assessments, which were
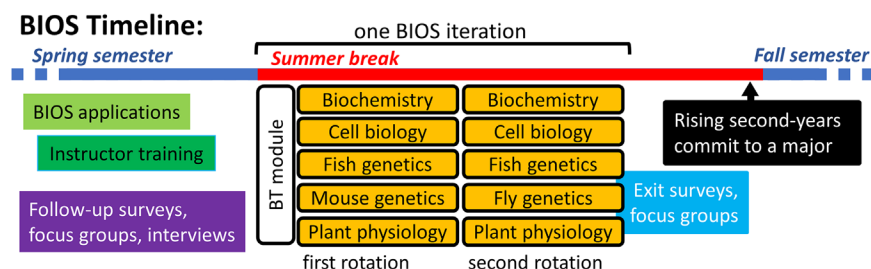
BIOS Timeline:

FIGURE 1. Timeline of BIOS events in the context of an academic year. Learner application (light green box) and instructor training (dark green) took place before each annual iteration of BIOS (black bracket), usually in the months of March through May. During each iteration (usually end of June to beginning of August), all undergraduate learners participated in a basic training (BT) module (white box) before participating in two topical tracks (orange boxes), each called a "rotation." At the end of each iteration, exit surveys and focus groups (blue box) were used to investigate program-level outcomes. Because most BIOS learners were rising second-year students, they committed to their undergraduate major after BIOS, before the start of the Fall semester (black box). In April and May of each subsequent year, follow-up surveys, focus groups, and interviews (purple) were used to investigate the long-term effects of BIOS enrollment on MR outcomes.

designed by an independent working group consisting of one or two faculty with expertise in that area and one or more of their graduate students and/or research staff. BIOS enrollees were allowed to participate in two back-to-back, each called a rotation (Figure 1).

Five activity types were incorporated into each track: lab, discussion, proposal, report, and workshop. Labs involved students doing experiments or preparations for experiments in a laboratory setting, while discussions were used to deliver and confirm the transfer of conceptual knowledge, often in a classroom setting. Proposals were used to give students the opportunity to make decisions about their experiments and predict the results. Reports were used to give students time to organize results, make figures, do statistics, and present their findings, usually in an open class setting. Finally, workshops were used to train students in research skills not requiring laboratory equipment and space, such as experimental design and statistics.

BIOS Participants

BIOS participants included faculty, postgraduates, and undergraduates (Table 1). Faculty managed or co-managed a topical track and were usually the ones who designed that track. Due to unpredictable faculty schedules, most day-to-day CURE instruction was handled by postgraduates. Faculty were strongly encouraged to interact as much as possible with undergraduate learners, especially during discussions, proposals, and reports. There was significant variation in this involvement between tracks and from year to year. In addition to track management, faculty assigned final grades, because BIOS was operated as a CURE awarding credit. A majority of faculty from 2015 to 2019 were male (Table 1), and all were affiliates of the same School of Life Sciences.

Postgraduates were either the graduate students or lab staff of faculty (Table 1), tasked with day-to-day operation of one topical track. Although faculty rarely changed from year to year, postgraduates changed regularly: on average, only 38% from any year returned to participate in a later year. Postgraduates

were usually responsible for conducting learning activities and assessments, especially the ones requiring direct observation. Because these activities required a great deal of time and energy, undergraduate teaching assistants were added in 2017 (Table 1) to help with assessments and lab preparations. These assistants were always students who had participated in BIOS before as learners.

From 2015 to 2019, more than 97% of all BIOS learners were rising second-year undergraduates. This demographic was specifically recruited for reasons described in the *Results*. A voluntary application process was used for recruitment, with applicants required to answer questions about future academic and career aspirations, such as: "How interested are you in doing science research?" and "Why do you want to join this program?" Applicants were made aware of the six topical tracks they could experience and required to rank-order their top four preferences. These preferences were used to assign each accepted student into two topical tracks.

Completed applications were reviewed by BIOS faculty, and selections were made based on two criteria: reasons for interest in research and reasons for wanting to experience a preferred track. Grades and other aspects of prior academic performance were excluded to remove bias in favor of high achievers. Faculty were strongly encouraged to recruit a mix of candidates with high and low self-reports of research interests. Students with low reports were targeted in an effort to use BIOS to help improve those opinions. Despite the attempt at balanced recruitment, BIOS applicants likely still disproportionately represented students with high research interests, an important consideration when interpreting the results (see *Discussion*). At enrollment, most learners reported intended majors in either the life or medical sciences (Table 1).

Exit Surveys

Activity-specific rubrics were used to collect empirical data on improvements in individual skills learned in each topical track (examples in *Results*). Program-level outcomes, however, were difficult to collect in this manner, motivating the use of surveys, focus groups, and interviews instead (Figure 1). Two types of exit surveys were administered to BIOS learners: track-specific and program-specific. Track-specific surveys were administered immediately after the conclusion of a topical track and included questions specific to that track. Program-specific surveys were administered at the end of BIOS and included broader questions about overall experiences.

In total, exit surveys contained 34 common questions used across all six topical tracks. Fourteen collected mentor-specific information to monitor instructional quality (e.g., "How much do you feel your postgraduate mentor made him/herself available for your questions?") and are not reported. The remaining 20 questions (see Supplemental Material), consisting of two types—pre–post and reflection—are. Pre–post questions were administered on the first and last days of the program, with the

TABLE 1. BIOS participant demographics, sorted by year of involvement (columns)

	2015	2016	2017	2018	2019
Faculty instructors					
Total	10	10	10	10	8
Female	2	3	3	3	3
Male	8	7	7	7	5
Postgraduate instructors					
Total	10	12	15	14	16
Female	6	7	8	8	8
Male	4	5	7	6	8
Grad student	7	10	11	10	11
Lab staff[a]	3	2	1	0	0
Undergraduate teaching assistants					
Total	0	0	3	4	5
Undergraduate learners					
Total	41	49	52	53	50
Female	19	21	32	34	28
Male	22	28	20	19	22
Life science[b]	29 (30)	28 (24)	26 (22)	27 (28)	26 (26)
Medical science[b]	10 (10)	14 (16)	22 (26)	15 (20)	20 (19)
Other science[b]	1 (1)	5 (6)	4 (4)	2 (5)	4 (3)
Not science[b]	0 (0)	2 (3)	0 (0)	0 (0)	0 (2)
BIOS exit assessment participants (undergraduates)					
Surveys	40	43	52	45	50
Focus groups	10	15	16	12	12

[a]Consisted of postgraduate technicians or lab managers.
[b]Intended major as reported on BIOS application; applicants could also respond "unsure," which is why the total of selections does not always equal "Total"; parentheses indicate students who actually joined that major after completing BIOS.

difference in responses used as an indicator of change. Reflection questions were used only at the end to prompt students to reflect on their experiences. Examples include: "How much did you learn in BIOS compared with your expectations before BIOS?"

The 20 questions reported in this work were used to collect four types of information: self-assessments of learning gains, research interests, science conceptualization, and perceptions about cooperative learning. In learning gain questions, students were asked to self-assess improvements in specific science process skills. One science conceptualization question required respondents to make word associations with "science," similar to prior work by Nakiboglu (2008) and Gulacar *et al.* (2015). Student inputs were thematically analyzed using the inductive coding method described in the "*Coding of Focus Group, Interview Transcripts*" subsection. The "order of association" method favored by Gulacar *et al.* was not implemented, because respondents in this study were only asked to associate a small number of words (three).

All survey questions were designed collectively by BIOS-participating faculty (content validation) before review by a panel of 12 life science undergraduates (face validation). Because all but one of the questions were designed to query subjective feelings about specific aspects of the BIOS experience and not meant to be assessments of latent constructs, methods such as Cronbach's alpha were not employed to determine interitem reliability. The lone exception was the "science" word-association question, which can be construed to invoke latent constructs. However, because this question was presented open-endedly, there was no established method to conduct validation.

Exit Focus Groups

Although student self-reports can be useful for characterizing a wide range of subjective qualities (Lopatto *et al.*, 2008; Cuthbert *et al.*, 2012; Shaffer *et al.*, 2014), recent work has indicated self-assessments of learning do not always correspond with empirical measures, with respondents sometimes overestimating or even underestimating their improvements (Eva *et al.*, 2004; Schiekirka *et al.*, 2013; Ziegler and Montplaisir, 2014). In this study, the application of more rigorous methods such as an established science process skills test (Burns *et al.*, 1985; Dirks and Cunningham, 2006; Feyzíoglu, 2012; Kramer *et al.*, 2018) was avoided, because the endeavor would have entailed the laborious process of translation and validation.

To compensate for self-report unreliability, all postgraduate mentors and a random selection of undergraduate learners were recruited into separate exit focus groups at the end of each BIOS iteration (Figure 1). These discussions were used to corroborate exit survey results and collect additional detail about BIOS experiences from both the learner and instructor perspectives. On average, 13 learners participated each year, representing roughly a quarter of each year's BIOS roster (Table 1). Discussions were conducted in mixes of English and Chinese, audio-recorded, and transcribed into English.

Follow-Up Surveys

To document differences in MR outcomes between BIOS and non-BIOS peers as they progressed through their undergraduate careers, a series of annual follow-up surveys were conducted (Figure 1). Only students with committed majors in the School of Life Sciences were invited to participate. These surveys contained 10 questions (see Supplemental Material), five requesting demographic information and the rest asking about different aspects of MR progress: for example, whether students had started their first experience and whether they had multiple experiences.

Because undergraduates were generally unfamiliar with the term, MR was described explicitly as "working in a university science lab doing authentic research." The term "authentic research" was defined explicitly as "real research that tries to answer a question we do not know the answer to," with examples given to flesh out the definition (see Supplemental Material). Follow-up surveys were designed using the same process employed for exit surveys: faculty collaboration (content validation) followed by student panel review (face validation). No questions invoked latent constructs, making it inappropriate to apply methods like Cronbach's alpha or confirmatory analysis.

On average, 93 undergraduates responded to the follow-up survey each year, representing between 23.1% and 36.1% of the total School of Life Sciences population (Supplemental Table 2). Each annual cohort consisted of both students who had completed BIOS (henceforth referred to as "ex-BIOS") and those who never participated (henceforth "non-BIOS"). Some non-BIOS students were already of second-year or higher academic rank when the first iteration of BIOS was offered in 2015, meaning they never had an opportunity to apply. These students were tracked separately as a "pre-BIOS" cohort (Supplemental Table 2) and treated as a pre-implementation baseline group.

Follow-Up Focus Groups, Interviews

In parallel with follow-up surveys, follow-up focus groups were conducted to collect additional details about BIOS's impact on MR (Figure 1). Both activities were conducted each year in April or May (Figure 1), meaning the earliest ex-BIOS students could participate was 10 or 11 months after their CURE experiences had ended. On average, a total of 26 undergraduates—comprising a mix of ex-BIOS, non-BIOS, and pre-BIOS students—participated each year, representing between 5.3% and 9.9% of the total School of Life Sciences population (Supplemental Table 3). Discussions were audio-recorded and transcribed into English in a de-identifying manner. Some focus group students were repeat contributors who had already participated in a prior year. These repeats comprised less than 9% of any annual cohort (Supplemental Table 3).

To investigate MR outcomes from the mentor's perspective, follow-up interviews (Figure 1) were conducted with faculty and postgraduates who had *not* participated in BIOS. All interviewees had or were mentoring at least one BIOS and one non-BIOS student. In total, eight faculty and 19 postgraduates were interviewed. Most faculty were male while most postgraduates were graduate students with three or fewer years of MR mentoring experience (Supplemental Table 4). Discussions were audio-recorded and transcribed into English in a de-identifying manner.

Coding of Focus Group, Interview Transcripts

All focus groups and interviews were conducted in a semistructured manner with a set of prepared questions used to initiate conversation and additional questions asked based on the flow of discussion. This format afforded the opportunity to delve deeper into individual perspectives. Within each set of interactions (pre-CURE, exit, or follow-up), relevant points made in earlier discussions were repeated by moderators in later discussions so as many participants as possible could comment on the same issues.

All focus group and interview transcripts were thematically analyzed using the inductive coding (Fereday and Muir-Cochrane, 2006) method described by Stuckey (2014). A group of researchers read each transcript and worked cooperatively to identify excerpts corresponding with different themes. Eight researchers analyzed pre-CURE transcripts, while 10 and seven, respectively, analyzed exit and follow-up transcripts. Each group's work was conducted at a different time and independently of the others.

Unique codes were used to label excerpts pertaining to specific themes. The name of each code was designed to summarize the essence of that theme using a short description that clearly distinguished it from others. Once each excerpt had a code, each set of excerpts categorized using the same code were reanalyzed by a pair of researchers to confirm the assignments seemed reasonable. During this second round of analysis, old codes were sometimes refined or split into new codes to better represent relevant excerpts. In rare cases when there was disagreement about whether an excerpt should be moved to another code, a third researcher was brought in to resolve that conflict. The coding process was deemed complete when all excerpts were assigned a code confirmed by at least one pair of researchers and author J.F. Student-selected words associated with "science" in exit surveys were analyzed by seven researchers using identical coding methods.

Quantitative Analysis of Focus Group, Interview Transcripts

In pre-CURE discussions, the frequencies of mention for individual MR challenges were tabulated from transcribed excerpts. The first time any participant mentioned experiencing or agreeing with the existence of a challenge was counted as one mention. Additional exposition about the same challenge by the same person was not counted. Authors quickly realized this method was not a good representation of overall opinions in focus groups, because many participants never verbally expressed their views. There were, for example, many instances when participants would nod in agreement without offering any verbal expression, a form of concurrence that failed to be captured on transcripts. The lone exception was when moderators used a show-of-hands method to ask focus group participants which MR challenge they thought was most influential (see *Results*). Interviews were unaffected by this issue, because they were always conducted individually.

To achieve a more accurate measure of sentiments in subsequent exit and follow-up focus groups, moderators consistently employed a show-of-hands method in those interactions, asking participants to agree or disagree with important statements made during discussion. Because this public display is susceptible to social desirability bias, agreement with some sensitive
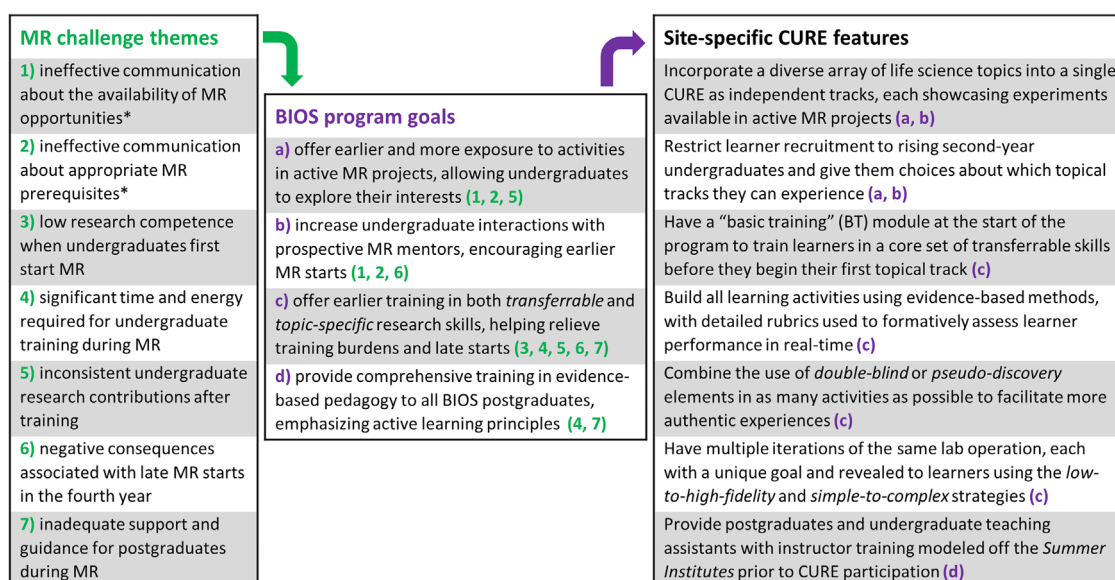
MR challenge themes	BIOS program goals	Site-specific CURE features
1) ineffective communication about the availability of MR opportunities*	**a)** offer earlier and more exposure to activities in active MR projects, allowing undergraduates to explore their interests **(1, 2, 5)**	Incorporate a diverse array of life science topics into a single CURE as independent tracks, each showcasing experiments available in active MR projects **(a, b)**
2) ineffective communication about appropriate MR prerequisites*	**b)** increase undergraduate interactions with prospective MR mentors, encouraging earlier MR starts **(1, 2, 6)**	Restrict learner recruitment to rising second-year undergraduates and give them choices about which topical tracks they can experience **(a, b)**
3) low research competence when undergraduates first start MR	**c)** offer earlier training in both *transferrable* and *topic-specific* research skills, helping relieve training burdens and late starts **(3, 4, 5, 6, 7)**	Have a "basic training" (BT) module at the start of the program to train learners in a core set of transferrable skills before they begin their first topical track **(c)**
4) significant time and energy required for undergraduate training during MR	**d)** provide comprehensive training in evidence-based pedagogy to all BIOS postgraduates, emphasizing active learning principles **(4, 7)**	Build all learning activities using evidence-based methods, with detailed rubrics used to formatively assess learner performance in real-time **(c)**
5) inconsistent undergraduate research contributions after training		Combine the use of *double-blind* or *pseudo-discovery* elements in as many activities as possible to facilitate more authentic experiences **(c)**
6) negative consequences associated with late MR starts in the fourth year		Have multiple iterations of the same lab operation, each with a unique goal and revealed to learners using the *low-to-high-fidelity* and *simple-to-complex* strategies **(c)**
7) inadequate support and guidance for postgraduates during MR		Provide postgraduates and undergraduate teaching assistants with instructor training modeled off the *Summer Institutes* prior to CURE participation **(d)**

FIGURE 2. Seven MR challenge themes (first column) motivated four BIOS program goals (second column) which in turn resulted in the design of seven site-specific features (third column). Asterisks (first column) denote themes describing events occurring before the start of a new MR experience; all others describe challenges arising during an MR experience. Green numbers (second column) correspond to MR challenge numbering, indicating which themes motivated each goal. Purple letters (third column) correspond to program goal lettering, indicating which goals motivated each site-specific feature.

statements (e.g., "Do you feel faculty exhibit gender bias that favors males?") was assessed via a slower online anonymous polling method using the Chinese app WeChat (Tencent Inc., Shenzhen, China). Data collected using either method are appropriately identified as such throughout this work.

Quantitative Analysis of Survey Results

For Likert-scale survey responses, both five- and four-point scales were used. To determine the statistical significance of response differences between cohorts or questions, Mann-Whitney U-test was performed in the R software v. 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) as described by Mangiafico (2015). Unlike a t test, Mann-Whitney U-test is nonparametric and does not assume either a normal or continuous distribution, making it especially appropriate for analyzing range-limited distributions with possible bi- or multimodality (Jamieson, 2004). For binary responses ("yes" or "no"), the $n - 1$ chi-square test was used, as recommended by Campbell (2007) and Richardson (2011). Confidence intervals for $n - 1$ tests were determined using the method recommended by Altman and colleagues (2000). An alpha level of 0.05 was employed for all tests, with Sidak-Bonferroni corrections applied when necessary. This study was considered exempt by IRB review.

RESULTS

Challenges Associated with Mandated MR

Pre-CURE focus groups and interviews revealed seven challenges associated with mandated MR. As shown in Figure 2, these themes were used to establish four BIOS program goals, which in turn, informed the design of seven site-specific features. Frequencies of mention for each challenge are summarized in Supplemental Table 5 with corresponding excerpts presented in Supplemental Table 6.

The first challenge was ineffective communication about the availability of MR opportunities. This theme was mentioned most often by undergraduates (Supplemental Table 5), who described various obstacles when trying to find available openings: these included a lack of resources to identify prospective mentors and a lack of guidance about how to approach them (Supplemental Table 6, excerpts 1a and b). On the mentor end, faculty sometimes mentioned difficulties promoting their MR projects, especially when trying to connect with prospective mentees (Supplemental Table 6, 1d). Both sets of concerns were described as increasing the likelihood that undergraduates would settle for MR projects they found less interesting because they lacked options when trying to find a mentor to satisfy the MR graduation requirement (Supplemental Table 6, 1a, b, and d).

The second challenge was ineffective communication about MR prerequisites. This was again mentioned most by undergraduates (Supplemental Table 5), who expressed concerns about not knowing what qualifications were needed to be considered a viable candidate. Many excerpts (Supplemental Table 6, 2a and b) demonstrated this lack of clarity was driving an underlying concern of rejection, which in turn was motivating many to wait until their third or fourth years before seeking their first opportunity, a delay many said, in retrospect, seemed unnecessary.

A third challenge was low research competence when first starting MR. Mentors mentioned this most often (Supplemental Table 5), with a postgraduate saying, "Clearly, upperclassmen do know a little more [than underclassmen] … But I find lab experience and basic lab operations are in lacking for everyone … Often, it is the simple things like adjusting the final volume after making a solution or using proper micropipetting techniques that they lack, things we assume they should have learned [by now]." Undergraduates generally corroborated this

lack of competence, with many opining in-semester lab courses did not seem to adequately prepare them for MR (Supplemental Table 6, 3a and b). These excerpts indicated lab courses in the study context were usually "cookbook" in nature, suggesting great potential for CURE implementation.

The fourth challenge was the need to invest significant time and energy in undergraduate training during MR. Postgraduates were most vocal (Supplemental Table 5), opining that such responsibilities typically fell to them, creating an unrealistic conflict of demands between mentoring and their own research work (Supplemental Table 6, 4a and b). Most undergraduates admitted being aware of this tension (Supplemental Table 5), repeatedly describing it as a strong deterrent to seeking assistance during MR (Supplemental Table 6, 4c and d). Some even described feeling forced to learn in an inefficient, self-reliant manner because faculty and postgraduates failed to provide consistent support and guidance (Supplemental Table 6, 4e and g). These excerpts indicated effective training was an issue very central to successful MR experiences. When asked to select the single most influential challenge (Supplemental Table 7), 60% of mentors (12 of 20) and 88% of undergraduates (22 of 25) chose this training burden.

The fifth challenge was inconsistent research contributions after training. This issue was voiced primarily by faculty (Supplemental Table 5), who lamented the fact that successful training did not necessarily guarantee meaningful returns in publishable work. A postgraduate said, "Even after undergraduates learn the experiments they need to know, I find their contributions can be very different. Some become very engaged … others are just trying to meet their graduation requirement … I think productive students are either very conscientious or really enjoy the research topic." Both faculty and postgraduates agreed the lack of topical interest seemed a common impediment to productivity (Supplemental Table 6, 5a–c), suggesting successful matching of student interests to MR topics may be key to nurturing favorable outcomes.

Inconsistent contributions were often mentioned in conjunction with another challenge: negative consequences associated with late starts. Many undergraduates opined that starting MR in the fourth year seemed to exacerbate in-lab tensions, because mentors often seemed to offer less support and guidance than for students who started earlier (Supplemental Table 6, 6a and b). Some postgraduates admitted to this bias, saying late starts seemed to significantly reduce the prospects for productivity, making their efforts feel less rewarding and diminishing their enthusiasm for active involvement (Supplemental Table 6, 6d–g). These excerpts suggested undergraduates who started MR earlier were more likely to be supported and, consequently, to be more productive. Most mentors agreed MR starts in the fourth year felt late, while those in the third did not (Supplemental Table 6, 6d, e, g, and h).

The final challenge was inadequate support and guidance for postgraduates during MR. In addition to experiencing unrealistic expectations for balancing mentoring with their own research work (Supplemental Table 6, 4a and b), most postgraduates agreed it was common 1) to have no prior teaching experience when mentoring someone for the first time and 2) to receive little or no support from faculty during mentoring (Supplemental Table 6, 7a and b). This lack of support and experience suggested an instructor training program might be a useful way to empower postgraduates to be more effective and independent mentors, an idea subsequently incorporated into BIOS as a program goal (Figure 2).

BIOS Program Goals

To help mitigate the identified MR challenges, four program goals were established. The first was to offer undergraduates earlier and more direct exposure to research in active MR projects. A goal motivated by three challenges (Figure 2a), it was hoped the incorporation of authentic activities into BIOS would help undergraduates feel a greater sense of access and opportunity to explore their interests before starting their first MR experience. Presumably, this would also help them construct their own firsthand understanding of authentic research, allaying concerns about prerequisites and ultimately empowering them to find research areas where they could be more engaged and productive.

The second goal was to provide BIOS learners the opportunity to interact with more prospective MR mentors. This too was a response to three challenges (Figure 2b). It was hoped undergraduates would build more meaningful research relationships if BIOS modules were taught exclusively by prospective mentors, helping motivate earlier MR starts. This approach to CURE instruction contrasts sharply with the employment of staff instructors without involvement in MR, increasing the potential for CURE experiences transitioning into MR. It was hoped that undergraduates would acquire more secondhand knowledge about authentic research through more contact with prospective mentors, again helping to allay concerns about prerequisites. Because pre-CURE focus groups had indicated undergraduates tended to interact more often and more comfortably with postgraduates than faculty (Supplemental Table 6, 1b and c), the recruitment of both into BIOS seemed a powerful way to encourage constructive dialogue.

The third goal, in many respects, was the most important: provide earlier undergraduate research training. A response to five MR challenges (Figure 2c), this goal motivated the design of a comprehensive skill-oriented CURE experience. From the onset, it was clear that there were, in fact, two sets of skills to consider. The first, *transferable skills*, applied broadly to most life science projects, regardless of topic, including activities like micropipetting and solution making. The second, *topic-specific skills*, applied only to discrete areas, including things like mouse tail cutting and fly sexing. Providing a balance of both through BIOS was intended to ensure program enrollees would require less training when starting their first MR experience.

Related to undergraduate training was the issue of instructor training. Pre-BIOS interviews had revealed postgraduates usually began MR mentorship with little or no teaching experience and minimal guidance about how to train undergraduates (Supplemental Table 6, 7a and b). Because the employment of these individuals as BIOS instructors ran the risk of precipitating poor outcomes, the fourth BIOS goal was to provide all postgraduates with comprehensive training in evidence-based pedagogy (Figure 2d). This training was designed to expose them to nondidactic principles like active learning and problem-based learning (Savery, 2015; Owens *et al.*, 2018), both core components of CUREs (Auchincloss *et al.*, 2014; Corwin *et al.*, 2018). Not only did a standardized training process offer the potential for immediate benefits to BIOS instruction, it also

offered the long-term potential to improve downstream MR outcomes, reducing the need for faculty guidance when BIOS postgraduates engaged in MR after BIOS.

It was immediately evident that these goals were not achievable within the span of a typical in-semester course or workshop. BIOS was, therefore, designed as a highly intensive 6-week CURE that acclimated and oriented undergraduates to the rigors of authentic research, an emphasis captured in the program's name: Biology Intensive Orientation Summer. BIOS was designed with full-day schedules, including weekends, giving time for learners to experience two topical tracks back-to-back, each running for 16 consecutive days (Figure 1). Participants experienced an average of 106 hours of learning activities per track, allowing many tasks to be linked together into dependent chains, as described in the following section.

Site-Specific CURE Features

Site-specific features were BIOS design elements added on top of the typical CURE framework as interventions to specific MR challenges (Figure 2). As described in the *Results* and *Discussion*, most of these features were later found to encourage favorable MR outcomes.

The first feature was the incorporation of six diverse life science topics into a single CURE, each implemented as an independent track containing experiments from active MR projects. Although the use of MR-derived activities in a CURE is not novel (Walcott *et al.*, 2018; Cole *et al.*, 2021), the level of diversity attained in BIOS is a significant departure from most prior designs, which have typically focused on only one or a small number of related topics, with learners seldom given choices about what they will experience (Kloser *et al.*, 2011; Brownell *et al.*, 2013; Brownell and Kloser, 2015; Corwin *et al.*, 2015; Ballen *et al.*, 2017). The diversity offered through BIOS gave learners a greater opportunity to explore their interests and interact with different prospective mentors, in line with the first two BIOS goals (Figure 2).

The second feature was a learner recruitment process that restricted BIOS application to rising second-year undergraduates. Although some comments in pre-CURE discussions had suggested these students might be too inexperienced to participate in intensive research (Supplemental Table 6, 6h), the fact that a plurality were already starting their first MR experience in their third year (unpublished data) motivated the decision to focus on a cohort at least a year younger, maximizing the potential benefits of early training and topic exposure (Figure 2).

The third feature was a 3-day "basic training" (BT) module that all BIOS learners were required to complete before entry into the first topical track (Figure 1). The BT module focused on PCR (polymerase chain reaction) and was used to train a core set of transferable skills. PCR was chosen, in part, because it was the task most prevalent in topical tracks. Given that undergraduates entered BIOS with varying degrees of high school and first-year lab experience, it was important to normalize these competences so students could start topical work on more equal footing. To this effect, the BT module was operated with the assumption that some participants had no prior biology lab experience at all. A series of competence tests was used to establish the pre-learning abilities of each student, giving the opportunity for those with high competence to test out. To be eligible, students needed to demonstrate basic PCR knowledge

and perform a successful reaction independently without mistakes. On average, only 7% of all participants from 2015 to 2019 were able to test out. A future publication will describe the BT module in greater detail, reporting the learning activities, competence tests, and outcomes for each annual cohort. Most students, regardless of year, spoke very highly of the BT module as a critical acclimation experience that helped them perform better in topical tracks.

The fourth feature was the deployment of evidence-based learning methods in all BIOS activities, part of a systematic effort to implement modern techniques in line with recent STEM education trends (Freeman *et al.*, 2014; Suchman, 2014; Bathgate *et al.*, 2019; Felege and Ralph, 2019). Although these methods are recommended, they are not a defining feature of CUREs, which can contain non–evidence-based practices like lectures. In BIOS, group discussions were used instead of lectures to deliver conceptual knowledge, while problem-based learning (Keller, 2002; Allchin, 2013) was implemented whenever possible in critical-thinking activities. In-lab student performance was monitored using detailed rubrics, each employed at least twice to give "pre-learning" and "post-learning" assessments, as recommended by Brownell and Kloser (2015). In most cases, only the final assessment was summative, with all others used formatively to inform learners about their progress in real time. Supplemental Figure 1 shows a rubric used to assess micropipetting performance in the BT module. As illustrated in this example, most rubrics recorded two elements of performance: 1) *process*, which required instructors to observe students while they operated lab instruments; and 2) *result*, which required evaluating operational outcome. Process assessments required a lot of firsthand observation, necessitating a learner to instructor ratio no greater than four. Track-specific learning gains determined through rubrics are beyond the scope of this work but will be the focus of future publications.

The fifth feature was the combined use of two methods to deploy discovery elements in learning activities: *double-blind* and *pseudo-discovery*. Double-blind discovery describes an experiment in which neither students nor instructors know the outcome, the type most often associated with CUREs (Corwin *et al.*, 2018; Goodwin *et al.*, 2021a). Pseudo-discovery, on the other hand, describes experiments in which learners are tasked with designing and performing experiments to discover outcomes instructors already know but refrain from revealing. This method was implemented so students could learn established ideas in a way that stimulated exploratory enjoyment. In the BT module, for example, a pseudo-discovery activity required students to choose different agarose gel concentrations when designing an experiment to reveal how that parameter affects DNA migration. To protect the integrity of the discovery element, students were told in advance what pseudo-discovery activities are and how they are meant to mimic authentic critical-thinking exercises. This explanation was accompanied by an explicit request to resist the temptation to use the Internet or other resources to learn outcomes prematurely. Through in-class observations and student feedback, compliance with this request was found to be extremely high. Distinctions between pseudo-discovery, student-defined discovery, and inquiry are addressed in the *Discussion*.

The sixth feature was the purposeful scheduling of multiple iterations of most lab activities, each assigned a unique goal
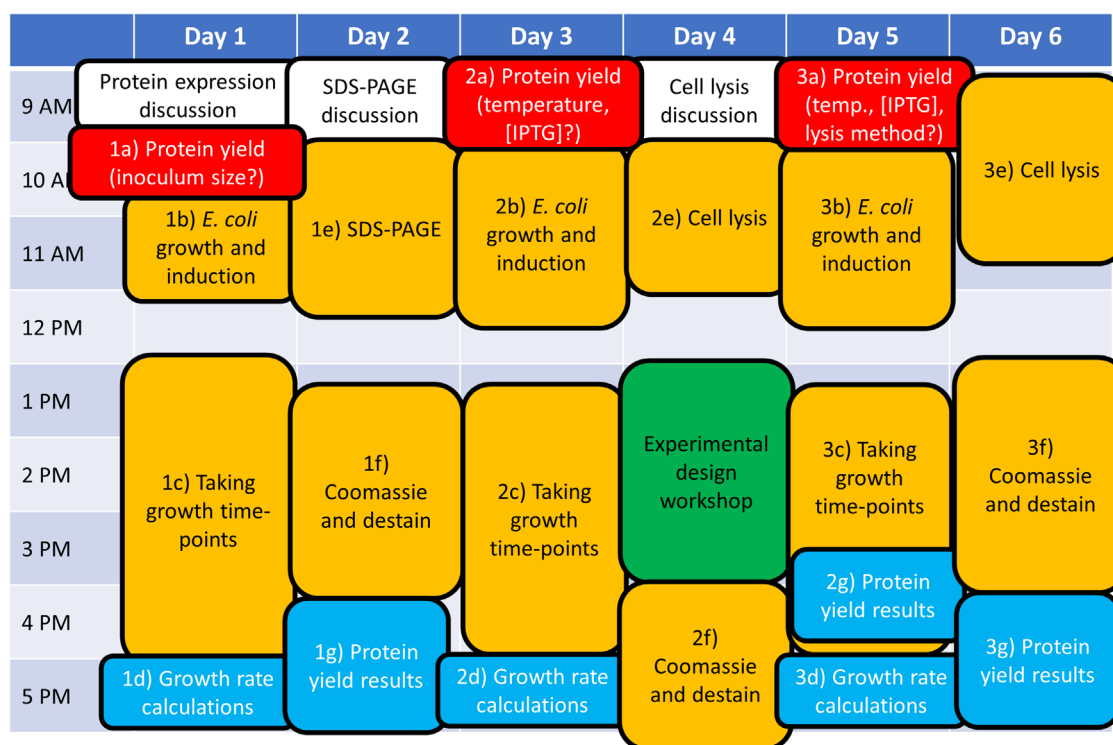
FIGURE 3. **Example of a learning activity schedule from BIOS's biochemistry track. As described in the** *Methods*, **each topical track contained five learning activity types: labs (orange bubbles), discussions (white), proposals (red), reports (blue), and workshops (green). Numbers indicate activities conducted in sequence using the same input materials. In series 1, students wrote a proposal predicting how inoculum size affects protein yield (1a), grew** *E. coli* **to test their prediction (1b), took time points to observe the growth curve (1c), and then calculated the growth rate (1d), recording their findings in notebooks. The next day, the same cells grown in 1b were used to conduct SDS–PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) and Coomassie staining (1e and f, respectively), with protein yields recorded in notebooks (1g). Once students became proficient at individual tasks, some were conducted in parallel, like 3c and 2g on day 5.**

(Figure 2). Although iterations are a defining feature of CUREs (Bakshi *et al.*, 2016; Ballen *et al.*, 2017), there is no established convention for how they should be implemented. In BIOS, iterations of the same activity were presented with unique objectives to make them feel as intellectually stimulating as possible. In the biochemistry track, for example, participants grew *E. coli* and conducted protein expression using an IPTG (isopropyl β-D-1-thiogalactopyranoside) induction three times over the course of 5 days (Figure 3). Each iteration was preceded by a proposal in which students were asked to consider how certain variables might affect protein yield. The first required students to predict the effects of starting inoculum size (Figure 3, bubble 1a), while the second and third, respectively, required the added consideration of IPTG concentration and lysis method (Figure 3, 2a and 3a). This process of incrementally increasing the variables to consider is referred to in cognitive load theory as the *low-to-high-fidelity* strategy (Van Merriënboer and Sweller, 2010), a method for managing task difficulty so students do not feel overwhelmed by new information. The *simple-to-complex* strategy (Van Merriënboer and Sweller, 2010) was employed in parallel, with students first being presented isolated elements of a task before gradually working up to its full complexity. In the biochemistry track, this principle was applied by first providing students with *E. coli* starter cultures and later having them pre-

pare their own. In this manner, students gradually took more and more ownership over each task, learning the intricacies of a complete operation through a choreographed sequence of revealed elements.

The final feature was evidence-based instructor training for all BIOS postgraduates (Figure 2). Training activities were modeled on workshops used in the Summer Institutes on Scientific Teaching (Pfund *et al.*, 2009; Couch *et al.*, 2015; Fendos, 2020b;) and implemented over a 3-week, 15-day period before BIOS (Figure 1). The first week focused on introducing a foundation of active-learning and scientific teaching concepts through open-ended discussions and presentations. The second focused on giving trainees the opportunity to build their own active-learning modules, while the third focused on learning the modules each trainee was scheduled to teach during BIOS. This process allowed participants to improve their instructional skills while also developing a greater understanding of the reasons why BIOS was designed using nondidactic methods. Trainees generally responded very positively to the experience, although most also admitted, after completing their first iteration of BIOS, that the training alone did not prepare them for all aspects of non–lecture style instruction. Details about instructor training and associated outcomes are beyond the scope of this piece but will be the focus of a future article.

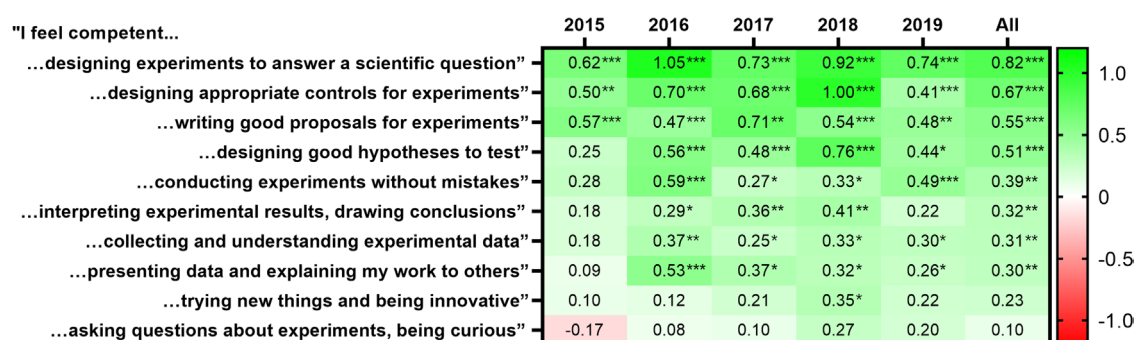| "I feel competent...	2015	2016	2017	2018	2019	All
...designing experiments to answer a scientific question"	0.62***	1.05***	0.73***	0.92***	0.74***	0.82***
...designing appropriate controls for experiments"	0.50**	0.70***	0.68***	1.00***	0.41***	0.67***
...writing good proposals for experiments"	0.57***	0.47***	0.71**	0.54***	0.48**	0.55***
...designing good hypotheses to test"	0.25	0.56***	0.48***	0.76***	0.44*	0.51***
...conducting experiments without mistakes"	0.28	0.59***	0.27*	0.33*	0.49***	0.39**
...interpreting experimental results, drawing conclusions"	0.18	0.29*	0.36**	0.41**	0.22	0.32**
...collecting and understanding experimental data"	0.18	0.37**	0.25*	0.33*	0.30*	0.31**
...presenting data and explaining my work to others"	0.09	0.53***	0.37*	0.32*	0.26*	0.30**
...trying new things and being innovative"	0.10	0.12	0.21	0.35*	0.22	0.23
...asking questions about experiments, being curious"	-0.17	0.08	0.10	0.27	0.20	0.10

FIGURE 4. Self-assessed BIOS learner skill improvements reported through exit survey pre–post questions on a four-point Likert scale (4 = agree, 3 = agree a little, 2 = disagree a little, 1 = disagree). Values are the average of differences between pre and post responses (post minus pre) for each annual cohort (columns). The "All" column depicts average differences for all cohorts combined. Asterisks indicate statistical significance, determined by comparing pre and post response distributions using Mann-Whitney *U*-test: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; all response distributions are provided in Supplemental Figure 2. Skill types (rows) arranged top-down in descending order of "All" column values.

Exit Surveys and Focus Groups

As described in the *Methods*, program-level CURE outcomes were quantified using exit surveys (Figure 1), with four types of information collected: self-assessed learning gains, research interests, science conceptualization, and perceptions about cooperative learning. These surveys were accompanied by exit focus groups, which added contextual detail.

Learners Reported Many Learning Gains. Through pre–post survey questions, BIOS learners reported a wide range of statistically significant learning gains. The largest and most consistent were for experimental design, control design, and proposal writing (Figure 4). Another four skills—designing hypotheses, conducting experiments, collecting data, and presenting data— were reported with significant improvements in four of five BIOS iterations. Only two—being innovative and asking good questions—were reported with improvements in only one or no iterations (Figure 4).

Interestingly, "being innovative" and "asking good questions" were reported with the highest average pre-program self-assessments (Supplemental Figure 2i and j), while "designing appropriate controls" and "writing good proposals" were the lowest reported (Supplemental Figure 2b and c). This indicated different skills had varying starting potentials for improvement, with BIOS helping to enhance areas with high self-perceived weakness. Focus group undergraduates cited activity iterations and their frequent variations in objectives as key reasons for the perceived gains (Supplemental Table 8, 1a–c), with one saying, "Even if it was the same experiment, having a different objective each time allowed us to practice and think about each task differently … I think this was very effective in helping me master each experiment."

Responses to post-program reflection questions were similarly favorable. Most students reported having learned "a lot" and more than they expected before BIOS (Table 2). Many focus group undergraduates said BIOS was the most satisfying research experience they ever had, with a variety of reasons accompanying this declaration (Supplemental Table 8, 1d–g). Some cited BIOS's authentic nature as a key contributor, while others mentioned its many formative assessments. Regarding the latter, one student said, "In BIOS, I felt like the learning was

tailored to students with a focus on the process, not just getting the right answer … Mistakes were not penalized the same way as in other classes so I felt more comfortable asking questions and really exploring each experiment." Double-blind and pseudo-discovery elements were also frequently complimented as important contributors to learning and engagement (Supplemental Table 8, 1h and i).

When asked to choose three things they thought were the most valuable BIOS outcomes, 90.6% of students selected "improve[d] my lab skills," while 66.9% and 40.8%, respectively, selected "hav[ing] fun doing experiments" and "gain[ing] more science knowledge" (Supplemental Table 9). In addition to reporting skill gains, many undergraduates were vocal about improvements in research confidence (Supplemental Table 8, 1j and k). When asked to vote by show of hands, 60 of 65 exit focus group participants said they felt their research confidence had improved through BIOS. Fifty-two said they felt ready to start their first MR experience, indicating successful realization of the second BIOS goal (Figure 2).

Research Interests Improved Despite Topical Interests Sometimes Declining. Three reflection questions were administered at the end of each topical track to determine research interest changes. The first and second asked, respectively, how much students were interested in "continuing research in the same area as your topical track" or "joining the lab of a professor who mentored in this track." The third asked students to rate whether their "interest in this track's area of research has increased or decreased." Overall, students reported increases in interest across all tracks and moderate to high interest in continuing work in the same lab or area (Table 3). This suggested initially that BIOS was a consistent positive influence on topical interests.

Focus group opinions, however, deviated noticeably. When asked to vote by show of hands, 25 of 65 participants reported their interest in one of the two tracks they took had actually decreased. This was quite different from the 10% reported through surveys (Supplemental Figure 3). When asked about the discrepancy, many admitted being hesitant to report negative feedback on surveys, because they did not want their instructors to feel bad or get in trouble. Thirty of 65 admitted favorably exaggerating their responses in some manner, with

TABLE 2. Self-assessed BIOS learning outcomes reported through exit survey reflection questions[a]

"How much …	2015	2016	2017	2018	2019	All
…do you feel you learned in BIOS overall?"	3.60 ± 0.99	3.74 ± 0.94	3.70 ± 0.85	3.79 ± 0.97	3.55 ± 1.02	3.70 ± 0.79
…do you think your lab skills improved?"	3.58 ± 1.10	3.69 ± 1.07	3.70 ± 0.90	3.65 ± 0.92	3.47 ± 0.92	3.64 ± 0.95
…fun did you have during BIOS?"	3.49 ± 1.05	3.65 ± 0.88	3.62 ± 0.89	3.60 ± 0.90	3.48 ± 0.97	3.60 ± 0.92
…did your overall science knowledge improve?	3.20 ± 0.98	3.62 ± 0.91	3.42 ± 0.80	3.44 ± 0.89	3.33 ± 1.06	3.42 ± 0.92
…did your understanding of the scientific process improve?"	3.20 ± 1.03	3.38 ± 0.90	3.30 ± 0.77	3.51 ± 1.01	3.27 ± 0.85	3.35 ± 0.77
…did your critical-thinking skills improve?"	2.90 ± 1.04	3.16 ± 1.03	3.02 ± 0.92	3.24 ± 0.99	2.90 ± 0.90	3.06 ± 0.80
…did you learn compared with your expectations before BIOS?"[b]	3.62 ± 1.62	3.90 ± 1.52	3.96 ± 1.49	4.07 ± 1.45	3.57 ± 1.39	3.84 ± 1.40

[a]Questions administered using a four-point Likert scale (4 = a lot, 3 = some, 2 = a little, 1 = none); all values are Likert score averages ± SD.
[b]Employed a five-point Likert scale (5 = a lot more than expected, 4 = a little more, 3 = same as expected, 2 = little less than expected, 1 = lot less).

most attributing the decline to deviations from pre-BIOS expectations. One said, "Before BIOS, I thought the experiments in [track name] would be colorful and exciting… Much to my surprise, I found the actual work was quite tedious." This sentiment was shared by many, suggesting experimental methods might be just as important as research topic for stimulating interest.

Despite the negative feedback, 58 of 65 focus group undergraduates agreed, when asked to vote anonymously via a mobile app (see *Methods*), that BIOS had improved their overall interest in life sciences research, even though it might have decreased their interest in one or more specific areas. Most opined the latter was actually a good thing, with one saying, "I was actually very grateful to get to experience [track name] and realize I don't like it … I think it helps a lot to know early on what topics you are not interested in so you can pick something better for your [thesis research]." These and other excerpts (Supplemental Table 8, 2c–f) again suggested the goal of early topic exposure was being successfully realized, with participants gaining a greater understanding of both authentic research and their own preferences. Students vehemently rejected the idea that their responses to other exit survey questions were exaggerated, allaying concerns that those responses might also be unreliable.

Science Conceptualization Changes Reflected the Fact That Research Is Difficult. Two pre–post questions were used to investigate BIOS's effects on science conceptualization. The first asked students to choose, from a predetermined list of science

process skills, the five they felt were most important (Figure 5). The second asked respondents to select three words to describe "science" (Figure 6). In pre-BIOS responses, learners rated "designing appropriate controls," "asking good questions," and "being innovative" as most important (Supplemental Table 10). Word associations, on the other hand, were found to favor two categories: intellectual appeal (e.g., "interesting," "fascinating," "puzzle") and emotional appeal ("beautiful," "colorful," "free"; Supplemental Table 11).

When pre–post responses were compared for each annual cohort, student inputs generally did not change in a statistically significant manner. However, when all cohorts were combined, some small but interesting patterns emerged. First, there was a statistically significant increase in the perceived importance of "conducting experiments without mistakes" and corresponding decreases in "designing good hypotheses to test" and "trying new things and being innovative." These changed in prevalence by 11.3%, 13.0%, and 18.8%, respectively (Figure 5). At the same time, words encapsulating the challenging nature of science (e.g., "arduous," "expensive," "difficult") increased by 25.5% while words emphasizing practical value ("useful," "necessary," "innovative") and discovery ("mystery," explore") decreased by 12.6% and 17.2%, respectively (Figure 6).

When asked about these changes, exit focus group undergraduates generally opined that they were reflections of BIOS experiences. One said, "I think, before BIOS, many of us didn't realize how difficult science can be and how slow it is in creating

TABLE 3. Self-assessed research interest changes reported through exit survey reflection questions[a]

Interested in…	2015	2016	2017	2018	2019	All
…continuing research in same area as your topical track?	2.99 ± 1.10	3.29 ± 1.05	3.08 ± 0.94	3.00 ± 0.97	2.86 ± 1.07	3.07 ± 0.77
…joining the lab of a professor who mentored in this track?	2.84 ± 1.13	3.31 ± 1.03	3.10 ± 1.02	3.07 ± 1.15	2.90 ± 0.99	3.05 ± 0.96
Interest in track's area of research has increased or decreased?[b]	3.78 ± 1.63	4.10 ± 1.60	4.02 ± 1.54	3.91 ± 1.50	3.72 ± 1.55	3.92 ± 1.39

[a]Administered using a four-point scale (4 = very, 3 = somewhat, 2 = a little, 1 = none); all values are averages ± SD; all response distributions in Supplemental Figure 3.
[b]Employed a five-point Likert scale (5 = increased a lot, 4 = increased a little, 3 = no change, 2 = decreased a little, 1 = decreased a lot).
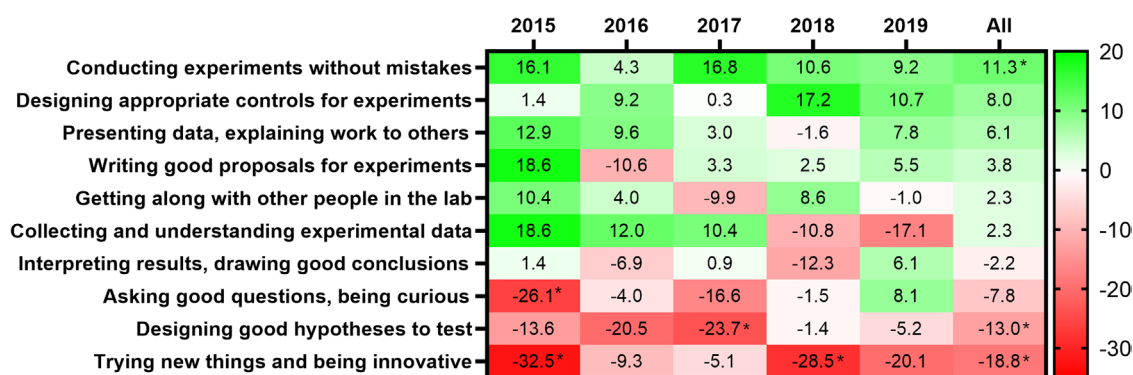
	2015	2016	2017	2018	2019	All
Conducting experiments without mistakes	16.1	4.3	16.8	10.6	9.2	11.3*
Designing appropriate controls for experiments	1.4	9.2	0.3	17.2	10.7	8.0
Presenting data, explaining work to others	12.9	9.6	3.0	-1.6	7.8	6.1
Writing good proposals for experiments	18.6	-10.6	3.3	2.5	5.5	3.8
Getting along with other people in the lab	10.4	4.0	-9.9	8.6	-1.0	2.3
Collecting and understanding experimental data	18.6	12.0	10.4	-10.8	-17.1	2.3
Interpreting results, drawing good conclusions	1.4	-6.9	0.9	-12.3	6.1	-2.2
Asking good questions, being curious	-26.1*	-4.0	-16.6	-1.5	8.1	-7.8
Designing good hypotheses to test	-13.6	-20.5	-23.7*	-1.4	-5.2	-13.0*
Trying new things and being innovative	-32.5*	-9.3	-5.1	-28.5*	-20.1	-18.8*

FIGURE 5. Science process skills (rows) BIOS learners thought were most important, determined through exit survey pre–post questions. Values are the difference in percent of students who selected each item when comparing pre and post responses (post minus pre) for each annual cohort (columns). Asterisks indicate statistical significance ($p < 0.05$), determined by comparing pre and post response distributions via Mann-Whitney *U*-test; all response distributions are provided in Supplemental Table 10. Rows arranged top-down in descending order of "All" column values.

results … I think I was naively under the impression that every experiment gives a new discovery so I never realized that most experiments are repeats or controls or even failed experiments … I think this might explain why many students may have started … associating science more with words like 'arduous' and 'difficult.'" These and similar comments (Supplemental Table 8, 3a–f) suggested BIOS learners had developed a more practical, operational understanding of science. As far as the authors are aware, this is the first time that CURE participation has been quantitatively linked to such a change.

Perceptions about Cooperative Learning Depended on Task Type. Because BIOS was designed with a lot of pair and group work, it was important to assess learner opinions about such activity. Three pre–post questions were used to investigate changes in perception by asking learners how they felt about "doing lab work in groups or pairs," "how useful" they felt such arrangements were, and "how comfortable" they felt "getting along with others in the lab." More than 90% of all learners began BIOS with positive views of both participation and usefulness (Supplemental Figure 4). When pre–post changes were analyzed for each annual cohort, there were very few statistically significant changes (Figure 7). When all cohorts were combined, there was a small but significant increase in comfort (Figure 7).

Focus group undergraduates offered nuanced explanations for these results, making frequent distinctions between the activities they thought were more or less useful when conducted cooperatively (Supplemental Table 8, 4a and b). One said, "When we were doing proposals or reports, I think it was very useful … to talk to each other and debate different ideas … When doing experiments, however, I think [cooperative learning] is maybe not so good. If someone I am working with makes a mistake, that will negatively affect my performance and my result." Aversion to the latter possibility was strong, even when students were asked to consider formative tasks (Supplemental Table 8, 4c). A few did mention potential benefits to cooperative activity when conducting complex experiments—such as the ability to catch one another's mistakes or help one another troubleshoot (Supplemental Table 8, 4d)—but these views were clearly in the minority, with 38 of 42 agreeing, by show of hands, that pair or group work seemed to offer many benefits for conceptual tasks and few for experimental ones. This opinion was consistent between annual cohorts and unlinked to topical tracks, suggesting a general relevance. As far as the authors are aware, this is the first time that CURE participation has been linked to such a task-specific divergence in opinion about cooperative learning.
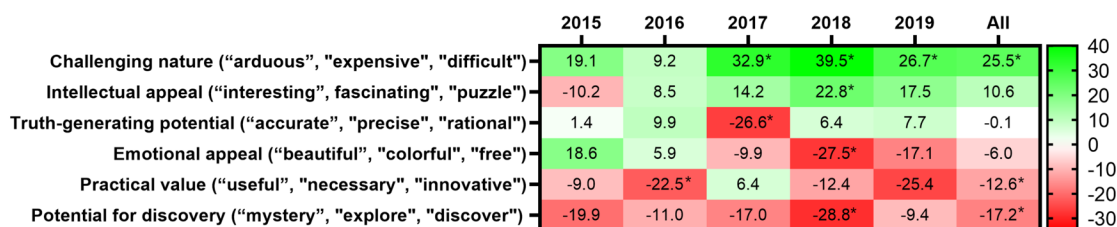
	2015	2016	2017	2018	2019	All
Challenging nature ("arduous", "expensive", "difficult")	19.1	9.2	32.9*	39.5*	26.7*	25.5*
Intellectual appeal ("interesting", fascinating", "puzzle")	-10.2	8.5	14.2	22.8*	17.5	10.6
Truth-generating potential ("accurate", "precise", "rational")	1.4	9.9	-26.6*	6.4	7.7	-0.1
Emotional appeal ("beautiful", "colorful", "free")	18.6	5.9	-9.9	-27.5*	-17.1	-6.0
Practical value ("useful", "necessary", "innovative")	-9.0	-22.5*	6.4	-12.4	-25.4	-12.6*
Potential for discovery ("mystery", "explore", "discover")	-19.9	-11.0	-17.0	-28.8*	-9.4	-17.2*

FIGURE 6. Thematically categorized responses when BIOS learners were asked to associate science with three words through an exit survey pre–post question. Values are the difference in percent of words falling into each category when comparing pre and post responses (post minus pre) for the same cohort (columns). Parentheses give examples of words falling into each category (rows). Asterisks indicate statistical significance ($p < 0.05$), determined by comparing pre and post response distributions via Mann-Whitney *U*-test; all distributions are provided in Supplemental Table 11. Word categories (rows) arranged top-down in descending order of "All" column values.
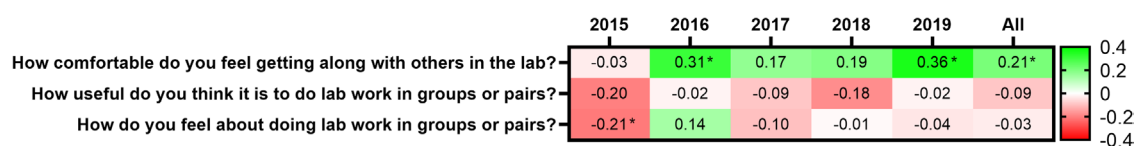
	2015	2016	2017	2018	2019	All
How comfortable do you feel getting along with others in the lab?	-0.03	0.31*	0.17	0.19	0.36*	0.21*
How useful do you think it is to do lab work in groups or pairs?	-0.20	-0.02	-0.09	-0.18	-0.02	-0.09
How do you feel about doing lab work in groups or pairs?	-0.21*	0.14	-0.10	-0.01	-0.04	-0.03

FIGURE 7. BIOS undergraduate perceptions about cooperative learning, as reported through exit survey pre–post questions on a four-point Likert scale (4 = very positive, 3 = a little positive, 2 = a little negative, 1 = very negative). Values are average differences between pre and post responses for each cohort (post minus pre). Asterisks indicate statistical significance ($p < 0.05$), determined by Mann-Whitney U-test; all distributions are provided in Supplemental Figure 4.

Follow-Up Surveys, Focus Groups, Interviews

Follow-up surveys were administered annually to track post-BIOS MR outcomes (Figure 1). One question asked whether respondents had started MR, while a second asked whether they had engaged in multiple experiences. Students who had not yet started MR answered extra questions about two barriers: the self-perceived need to prepare more before starting and whether it felt "difficult to approach professors about doing authentic research." Both had been identified in pre-CURE discussions as obstacles to early starts (Supplemental Table 6, 1b and 2a and b). Respondents were identified as non-BIOS or pre-BIOS based on year of matriculation, with the former being the direct peers of ex-BIOS and the latter being upperclassmen who never had the opportunity to apply. As described in the *Methods*, pre-BIOS students were treated as a pre-CURE baseline group, while non-BIOS were treated as a same-cohort control.

Ex-BIOS Students Started MR Earlier. Compared with pre-BIOS and non-BIOS students, ex-BIOS students reported a much higher rate of MR participation in their second and third academic years: over two times higher in the former and about 25% higher in the latter (Table 4). These differences were consistent for cohorts who matriculated in different years, suggesting BIOS was imparting a sustained influence on these outcomes. For respondents who said they had not yet started MR, ex-BIOS students were more likely to disagree with the idea that they needed to improve their science competence more (Table 5) and that it felt difficult to approach prospective faculty (Table 6). This indicated ex-BIOS students felt less hindered by these barriers, suggesting successful mitigation of corresponding MR challenges (Figure 2).

Focus group undergraduates overwhelmingly endorsed the idea that ex-BIOS students seemed to have greater confidence in their research skills, helping motivate earlier starts (Supplemental Table 12, excerpts 2a–c). Mentors, in contrast, offered less unified views. Some echoed undergraduate opinions about confidence (Supplemental Table 12, 2d), while others pointed to perceived differences in science knowledge or communicability: in the latter two, ex-BIOS students were generally perceived as having more than non-BIOS (Supplemental Table 12, 2e and f). Some non-BIOS students suggested ex-BIOS peers might enjoy an advantage in reputation when seeking MR opportunities because of their BIOS experience (Supplemental Table 12, 2g). When asked about this, only 10 postgraduates and six faculty agreed. Most opined that any advantage was likely small, far outweighed by the consideration of other factors like academic record, research interests, and personality.

More Ex-BIOS Students Participated in Multiple Mentored Experiences. Ex-BIOS students in their third or later years were more than twice as likely as pre-BIOS and non-BIOS students to report participation in multiple MR experiences (Table 7). Follow-up focus groups offered many circumstances viewed as helping to facilitate these experiences, such as the successful completion of a project or conflict with mentors (Supplemental Table 12, 3a and b). Both undergraduates and postgraduates agreed an early MR start often seemed a critical ingredient for multiple experiences (Supplemental Table 12, 3c and d).

TABLE 4. Percent of life science undergraduates reporting prior or ongoing participation in MR[a]

	Pre-BIOS					
Matriculated	2nd	3rd	4th+			
2011			90.0%			
2012		66.7%	100%			
2013	36.4%	75.0%	100%			
Averages[b]	36.4%[a]	70.6%[b]	97.1%[c]			

	Non-BIOS			BIOS		
Matriculated	2nd	3rd	4th+	2nd	3rd	4th+
2014	23.8%	60.0%	91.7%	90.0%	90.9%	100%
2015	45.5%	72.2%	100%	89.7%	100%	100%
2016	25.0%	68.8%		84.6%	94.7%	
2017	36.8%			93.1%		
Averages[b]	32.9%[a]	67.3%[b]	95.7%[c]	89.5%[c]	95.4%[c]	100%[c]

[a]"2nd," "3rd," and "4th+," respectively, refer to students in their second, third, and fourth or later year of academic study.
[b]Lowercase a–c indicate statistically distinct averages compared using the $n - 1$ chi-square test (i.e., an "a" and "b" indicate statistically distinct values) with Sidak-Bonferroni corrections.

TABLE 5. Undergraduate agreement with the idea that current science competence feels inadequate to start first MR experience[a]

	Pre-BIOS					
Matriculated	2nd	3rd	4th+			
2011			1.00			
2012		2.83 ± 1.70	NA			
2013	3.46 ± 1.39	2.80 ± 1.16	NA			
Averages[b]	3.46 ± 1.39[a]	2.82 ± 1.32[ab]	1.00			

	Non-BIOS			BIOS		
Matriculated	2nd	3rd	4th+	2nd	3rd	4th+
2014	3.37 ± 0.85	2.40 ± 1.10	2.00	1.67 ± 0.76	1.50	NA
2015	3.56 ± 0.22	2.50 ± 0.97	1.00	2.00 ± 1.30	NA	NA
2016	3.18 ± 0.98	2.60 ± 1.13		1.75 ± 1.28	1.00	
2017	3.31 ± 0.85			1.50		
Averages[b]	3.35 ± 0.72[a]	2.50 ± 1.14[b]	1.50	1.75 ± 1.17[c]	1.33 ± 0.28[c]	

[a]Each value is a Likert score average ± SD with responses coded on a five-point scale (5 = agree, 4 = agree a little, 3 = neither agree nor disagree, 2 = disagree a little, 1 = disagree).
[b]Lowercase a–c indicate statistically distinct averages compared using Mann-Whitney *U*-test with Sidak-Bonferroni corrections; values without error bars indicate cohorts of $n < 3$; "NA" indicates all students in that cohort had already started an MR experience; labeling corresponds to Table 4.

TABLE 6. Undergraduate agreement with the idea that faculty seem difficult to approach when looking to start first MR experience[a]

	Pre-BIOS					
Matriculated	2nd	3rd	4th+			
2011			2.00			
2012		3.18 ± 1.39	NA			
2013	4.34 ± 1.16	3.20 ± 1.00	NA			
Averages[b]	4.34 ± 1.16[a]	3.19 ± 1.37[b]	2.00			

	Non-BIOS			BIOS		
Matriculated	2nd	3rd	4th+	2nd	3rd	4th+
2014	4.05 ± 0.78	2.80 ± 1.09	4.00	3.33 ± 0.85	3.00	NA
2015	4.21 ± 0.85	3.09 ± 1.20	2.00	3.67 ± 1.31	NA	NA
2016	3.99 ± 0.81	3.00 ± 1.06		3.25 ± 1.18	4.00	
2017	4.08 ± 0.79			3.00		
Averages[b]	4.08 ± 0.70[a]	2.97 ± 1.05[b]	3.00	3.33 ± 1.39[b]	3.33 ± 0.58[b]	

[a]Each value is Likert score average ± SD with responses coded on a five-point scale (5 = agree, 4 = agree a little, 3 = neither agree nor disagree, 2 = disagree a little, 1 = disagree).
[b]Lowercase a and b indicate statistically distinct averages compared using Mann-Whitney *U*-test with Sidak-Bonferroni corrections; values without error bars indicate cohorts of $n < 3$; labeling corresponds to Table 4.

TABLE 7. Percent of life science undergraduates reporting multiple mentored experiences[a]

	Pre-BIOS					
Matriculated	2nd	3rd	4th+			
2011			10.0%			
2012		5.6%	15.4%			
2013	4.5%	6.3%	9.1%			
Averages[b]	4.5%[ab]	5.9%[b]	11.8%[b]			

	Non-BIOS			BIOS		
Matriculated	2nd	3rd	4th+	2nd	3rd	4th+
2014	0.0%	6.7%	16.7%	10.0%	22.7%	33.3%
2015	0.0%	5.6%	18.2%	6.9%	25.0%	40.0%
2016	5.0%	12.5%		7.7%	21.1%	
2017	0.0%			10.3%		
Averages[b]	1.2%[a]	8.2%[b]	17.4%[bc]	8.8%[b]	23.1%[c]	36.6%[d]

[a]Labeling corresponds to Table 4.
[b]Lowercase a–d indicate statistically distinct averages compared using the $n - 1$ chi-square test with Sidak-Bonferroni corrections.

For students who quit one lab and switched into another because of conflicts, confidence attributed to BIOS was often mentioned as a key factor enabling the move. One ex-BIOS student said, "I started [my first MR experience] right after BIOS … I soon found, however, that the project was not as fun or as rigorous as I expected so I quit and joined another lab … BIOS helped me understand what good research looks like … [It] also helped me feel very confident in my abilities … If I wasn't so confident or experienced, I think I would have just stayed in that [first] lab until I graduated because I would be afraid of not being able to find a second chance." This comment suggested ex-BIOS students were benefiting from a better understanding of authentic research, again pointing to mitigation of concerns with MR prerequisites. Most follow-up faculty and postgraduate interviewees felt unsure about whether ex-BIOS students were more likely to have multiple experiences, with only six of 27 saying they agreed with the idea.

Ex-BIOS Students Performed Better during MR. A surprising number of focus group and interview participants opined that ex-BIOS students seemed to perform better during MR. Many outcomes were cited, the most significant of which was research productivity (Supplemental Table 12, 4a). Seven of 14 postgraduates opined they felt ex-BIOS students were more productive. Voting by show of hands, 33 of 55 ex-BIOS and 20 of 41 non-BIOS undergraduates agreed. All others either felt unsure or thought there was no difference. A few suggested there were also differences in publication authorship (Supplemental Table 12, 4b), with ex-BIOS students viewed as achieving authorship more often. Because publication can be a sensitive topic among faculty, follow-up surveys did not evaluate this claim.

In addition to enhanced productivity, ex-BIOS students were suggested to be more self-sufficient during MR. By show of hands, 25 of 55 ex-BIOS and 17 of 41 non-BIOS students endorsed the idea, while 30 and 21, respectively, said they felt unsure or thought there was no difference. Many undergraduates gave examples of ex-BIOS students taking leadership or mentoring roles when interacting with other undergraduates in the same or an adjacent lab (Supplemental Table 12, 4c). This suggested the added experience and confidence obtained through BIOS might be helping facilitate beneficial lateral interactions with other MR mentees. Five of 14 postgraduates endorsed the idea that ex-BIOS students often seemed to lead other undergraduates. Faculty, on the other hand, were more cautious, with three agreeing ex-BIOS students might exhibit better performance and leadership, but only when their MR project was closely related to the skills learned in BIOS (Supplemental Table 12, 4e). This proposition is consistent with the constructivist framework inherent to CUREs (Cooper *et al.*, 2019), suggesting BIOS's influence on subsequent MR self-sufficiency might depend, at least partly, on CURE and MR task relatedness.

Other reasons given for better ex-BIOS performance during MR included science conceptualization and topical interest. Several ex-BIOS students suggested having a more realistic outlook on authentic research likely facilitated greater productivity and self-sufficiency, with one saying, "I think one of the important lessons that [undergraduates] need to learn about research is that it is difficult … I think the earlier you can learn this and overcome the shock, the sooner you can become productive and independent … I think BIOS helps us realize this very quickly." These comments suggested the observed changes in science conceptualization toward a more practical outlook (Figures 5 and 6) might be a beneficial development for fostering independence during MR. Topical interest, on the other hand, was mentioned most by faculty and postgraduates, who suggested ex-BIOS students often seemed surer of what they were interested in than non-BIOS peers, helping enhance engagement and productivity (Supplemental Table 12, 4f), an idea many undergraduates endorsed (Supplemental Table 12, 4g).

DISCUSSION

The present study describes a two-phase CURE implementation process in which MR challenges were identified and then targeted using site-specific features (Figure 2). A mixed methods approach was used to assess program outcomes, confirming CURE participants were more likely to engage in both earlier and multiple MR experiences (Tables 4, 7). Although empirical measures of individual skill improvements are not reported, focus groups and surveys nevertheless corroborated the fact that learning gains were systematically prevalent across all BIOS iterations (Figure 4 and Supplemental Table 8, excerpts 1a–l).

Given that the application of non-didactic learning methods in Asian contexts remains rare and is often viewed with skepticism (Fendos, 2018, 2020a, b), the use of an initial qualitative phase to characterize context-specific challenges offers important advantages. First, this information can be used to argue in favor of the need for reforms, combatting reticence about new learning implements (Fendos, 2020a, b). Second, it enables more informed selection of appropriate interventions, allowing each challenge to be targeted with specific program features (Figure 2). Third, these challenges can double as points of observation to determine intervention effectiveness, similar to the method described by Irby and colleagues (2018). In this study, for example, both surveys and focus groups employed questions about MR challenges, allowing observation of their mitigation (Tables 5 and 6). The use of overlapping qualitative and quantitative methods also allowed misleading survey responses to be identified (Table 3), highlighting an unexpected benefit of the mixed approach.

BIOS Features That Enhanced MR Outcomes

The results of this study indicate both common and site-specific CURE features played important roles in facilitating favorable MR outcomes. The use of task iterations and the authentic nature of those activities were common features mentioned many times as having strong positive effects on research competence and confidence (Supplemental Table 8, 1g–k), in line with prior work demonstrating similar outcomes in other contexts (Staub *et al.*, 2016; Peteroy-Kelly *et al.*, 2017; Ayella and Beck, 2018; Rodrigo-Peiris *et al.*, 2018; Murren *et al.*, 2019; Sewall *et al.*, 2020). In this study, both iterations and authenticity were frequently mentioned as important influences on science conceptualization, enabling BIOS learners to develop a more realistic appreciation of the difficulties involved in professional research. Most students opined that this appreciation was an important step in their development (Supplemental Table 8, 3a–e), agreeing with work by Goodwin and colleagues (2021a) in arguing that episodes of failure are a necessary and beneficial aspect of authentic experiences.

Among BIOS's site-specific features, three were mentioned most often as facilitating favorable MR outcomes: frequent formative assessments, topic diversity, and the targeted recruitment of rising second-year students. Formative assessments were mentioned most in exit focus groups, described as a positive contributor to research competence and confidence by helping students to understand their learning progress and correct their mistakes more effectively (Supplemental Table 8, 1e and i). BIOS topic diversity was mentioned most in follow-up discussions, described as an important way to let students discover their research interests, facilitating greater engagement and productivity during MR by enabling the selection of more compatible MR projects (Supplemental Table 12, 4f–g). Both mentors and mentees in follow-up discussions generally agreed the targeted recruitment of second-year undergraduates was an important facilitator of early training, helping encourage both early and multiple MR experiences (Supplemental Table 12, 1a–d and 3c and d).

Two other site-specific features were occasionally mentioned as facilitating favorable MR outcomes: 1) the use of double-blind and pseudo-discovery experiments and 2) the use of task iterations with unique objectives (Figure 2). Both were described as contributing to increases in research competence and confidence, the former by enhancing the authentic nature of BIOS activities (Supplemental Table 8, 1h and i) and the latter by helping students develop a more well-rounded understanding of the experiments they were learning (Supplemental Table 8, 1b and c). The low-to-high fidelity and simple-to-complex strategies associated with the latter were described by some as providing a useful stepwise structure with which to organize and construct new knowledge (Supplemental Table 8, 1d), a form of scaffolding known to be beneficial in other STEM contexts (Gross *et al.*, 2017; Fendos, 2021).

Site-Specific Features That Enhanced CURE Engagement

In addition to encouraging favorable MR outcomes, BIOS's frequent use of discovery elements also appears to have helped enhance CURE engagement. Both BIOS's double-blind and pseudo-discovery activities allowed learners to make experimental choices within the "sandbox" confines of a predetermined scientific question. This meant the activities were not pure inquiry (Moreira, 2013; Corwin *et al.*, 2018; Cooper *et al.*, 2019; Goodwin *et al.*, 2021a), aligning instead with the "student-defined" and "instructor-defined" definitions of discovery provided by Auchincloss *et al.* (2014). Activities in the fish genetics track, for example, allowed learners to make independent, student-defined choices (e.g., choose a chemical contaminant) within the confines of a larger, instructor-defined question (e.g., to see how it affects zebrafish embryo development). Because every topical track had some of these double-blind experiments, all tracks maintained the CURE requirement for general relevance. Having such ownership over their activities and the opportunity for discovery were frequently described by exit focus groups as feeling "fun," "like a game," and enhancing engagement (Supplemental Table 8, 1g–i).

Surprisingly, the regular use of formative assessments also seems to have enhanced student engagement. Many learners said the regular feedback helped them focus more effectively on the mistakes they were making, creating continuity between tasks. One student described this by saying, "The regular feedback was helpful because it gave you something to focus on the next day. If your [formative assessment] today said you made a mistake with [task name], you would naturally focus more on it tomorrow … This awareness of one's mistakes was very helpful in keeping us focused on the things relevant to our own learning." Another student described this focus on individual goals as "feeling like a game," with the next experiment or task being viewed as an opportunity to "level up" their research skills (Supplemental Table 8, 1e).

A third site-specific feature said to enhance CURE engagement was the use of the low-to-high fidelity and simple-to-complex strategies in task iterations (Figure 2). Despite having established utility in improving skill competence (Corwin *et al.*, 2018; Zydney *et al.*, 2020; Goodwin *et al.*, 2021a), iterations often run the risk of feeling boring. Revealing new tasks and variables in a choreographed manner through these two strategies made many BIOS learners feel like they were allowed to experience and consider different facets of the same experiment in a structured manner over a longer period, which helped to protect them from information overload while encouraging greater curiosity about the things being learned (Supplemental Table 8, 1b and d).

Context Matters for Cooperative Learning Preferences

An unexpected observation from this work is that learner perceptions of cooperative learning can be task-specific. Most BIOS learners agreed conceptual tasks like proposals and reports felt more useful to do cooperatively than experimental ones. Some acknowledged that group and pair work in the latter could help them avoid mistakes and improve troubleshooting, but these benefits often seemed outweighed by the potential dissatisfaction of having one's experiments ruined by someone else (Supplemental Table 8, 4a–c). This concern was expressed most strongly—but not exclusively—in the context of summative work, with students often opining that shared grades felt "unfair." This opinion motivated a separate study to examine which active-learning class operation rules students found fairer than others (Fendos, J. F., Cha, S. C., Yang, X. Y., Cai, L. C., and Yang, J. Y. unpublished data).

Because BIOS learners were affiliates of one of the most competitive universities in China, it is possible the dislike for shared grading was amplified by that competitiveness: with students being more sensitive to assessments of their work in general and, therefore, less likely to appreciate evaluations tied to others. The present study, unfortunately, does not provide decisive insight into this issue, only offering competitiveness as a possible influencing factor. Regardless of the underlying cause, the results advocate for a more cautious approach to cooperative learning implementation, one that better accounts for task-specific learner preferences.

Self-Sufficiency as Compensation for Mentor Absence

The observation that ex-BIOS students seemed more self-sufficient during MR was often mentioned in the context of mentor absence (Estrada *et al.*, 2018; Limeri *et al.*, 2019). In some follow-up excerpts, self-sufficiency was even described as necessary compensation for that absence, with ex-BIOS students repeatedly depicted as operating better under such conditions (Supplemental Table 12, 4a–c). Various accounts of ex-BIOS students taking on leadership and mentoring roles for other

mentees (Supplemental Table 12, 4b and c) illustrated the presence of lateral interactions in which BIOS students helped non-BIOS peers cope with their own struggles with mentor absence. This indicated, quite surprisingly, that BIOS may impart an indirect positive effect on students who never enrolled.

Similar lateral benefits were implied through comments about undergraduate teaching assistants. In exit focus groups, many learners complimented the assistants as a useful resource (Supplemental Table 8, 1l). Some even indicated the assistants could offer unique insights that postgraduate instructors could not, especially because the assistants had themselves already participated in BIOS as learners (see *Methods*). This potential for lateral benefit is seldom mentioned in work published in Western contexts. In work by Auchincloss *et al.* (2014), for example, self-sufficiency is often implied through ideas like "self-efficacy" and "resilience and grit," with no associated mention of how a subpopulation of students with high research competence may confer benefits to their peers. The present study offers evidence that such transfer through a CURE is possible.

Mandated MR as a Mixed Blessing

Because the present study was only conducted in a context with mandated MR, the results do not allow for a comparative analysis of how its absence may affect MR outcomes. When asked to reflect, faculty and postgraduates offered mixed views, with many saying the requirement felt like an unnecessary burden (Supplemental Table 6, 4a and b). Others said MR was an important educational service, with many postgraduates describing it as a meaningful way for them to "give back" for their own prior experiences (Supplemental Table 6, 7c and d). Despite divergent opinions, most mentors agreed the requirement was a significant source of in-lab tension, often harming postgraduate productivity.

Unlike mentors, undergraduates were more unified in their opinion that mandated MR was a good thing, guaranteeing access to authentic research training. One said, "If there is no [MR requirement], I think most professors would refuse to accept undergraduates." Faculty were divided in their agreement with this, with most saying they would likely still accept undergraduates but perhaps fewer of them. One said, "I think most faculty want [high-achievers] … So, perhaps the absence of the [requirement] would mean many [low-achievers] would find it difficult to [gain access]." Given that prior literature has demonstrated authentic experiences can often offer greater benefits for low achievers than high-achieving peers (Russell *et al.*, 2007; Thiry *et al.*, 2012; Brownell *et al.*, 2013), these comments would endorse the need to implement methods that alleviate the challenges associated with compulsory MR, in line with similar suggestions by Goodwin *et al.* (2021b). The results of this study would indicate an MR preparatory program like BIOS may be a useful way to provide such assistance, helping relieve specific burdens felt by MR mentors.

Caveats to BIOS Implementation

The unique nature of the study context does provoke the need to consider how well the BIOS design may translate into other locales. It is possible that BIOS learners, being among the most competitive in China, were especially well-positioned to leverage their experiences into favorable outcomes. A second consideration is the fact that BIOS participants were affiliates of a School of Life Sciences that enjoyed a very high ratio of roughly 0.4 faculty per undergraduate. This is certain to have offered learners a wider range of MR opportunities, strongly influencing their ability to engage in multiple experiences. This abundance also no doubt made it easier to recruit faculty from many research areas into a single CURE, a luxury many universities will not enjoy.

A third consideration is the availability of postgraduates. Compared with other universities in Asia, the present study context would be on the higher end for graduate student and lab staff availability, meaning a corresponding lower reliance on undergraduates for research productivity. It is unclear how this difference may have affected BIOS outcomes, but at the very least, universities with no graduate students would find a BIOS-modeled CURE difficult to implement because postgraduates are the primary instructional workforce. These locales would need to chart an alternative path to instruction that carefully considers how faculty might be encouraged or incentivized to take on more active roles.

Study Limitations

The present study's most significant methodological limitation is the BIOS learner recruitment process, which, by virtue of its voluntary nature, likely biased application in favor of students who, on average, already had greater interests in experimental research. It is possible this may have diverted students with greater interests into BIOS while leaving less-interested peers to accumulate in non-BIOS cohorts. Nothing in the outcome data can be used to refute this. At the same time, it was reassuring to see both pre-BIOS and non-BIOS cohorts behave almost identically in the timing of their first MR experiences and the prevalence of multiple experiences (Tables 4 and 7). This suggests, as least for these two outcomes, that BIOS may be an influence independent of pre-enrollment interest differences. Had research interests been the primary driver of outcomes, one would have expected non-BIOS cohorts to exhibit even later starts and even fewer multiple experiences than pre-BIOS peers.

Although the BIOS application process specifically excluded consideration of prior academic performance, a retroactive analysis determined BIOS students were more likely to have higher cumulative grade point averages (GPAs) than non-BIOS peers, both at the time of BIOS enrollment and at graduation (Supplemental Table 13). This analysis was only conducted for cohorts who matriculated in 2014, 2015, and 2016 to avoid complications associated with the COVID-19 pandemic. At-enrollment GPA differences between BIOS and non-BIOS peers matriculated in the same year were not large, averaging 0.32 on a four-point scale. This nevertheless offers the possibility that some uncharacterized difference in student aptitude may be contributing to the MR outcome differences observed between groups. Interestingly, at–BIOS enrollment and at-graduation GPAs were not statistically different for BIOS learners (Supplemental Table 13), indicating BIOS had no effect on in-semester course performance. Scientific identity, a common focus of other publications (Hurtado *et al.*, 2011; Robnett *et al.*, 2015; Goodwin *et al.*, 2021a), was not examined, leaving its relevance unknown.

CONCLUSIONS

In closing, the present study offers a compelling model for a CURE implementation process that can improve downstream

MR outcomes. The BIOS design showcases site-specific features that can help mitigate common challenges associated with compulsory MR, challenges present across the globe in many locales. As interest and activity in CURE design continues to increase and mature internationally, it is hoped this work can make an important contribution by offering new ways to consider the importance of context in CURE implementation, especially when dealing with longitudinal program goals that extend beyond the typical life span of a CURE. For many faculty who depend on MR for research productivity, this work should be of considerable interest.

REFERENCES

Abu-Zaid, A., & Alkattan, K. (2013). Integration of scientific research training into undergraduate medical education: A reminder call. *Medical Education Online*, *18*(1), 16–19. doi: 10.3402/meo.v18i0.22832

Aikens, M. L., Sadselia, S., Watkins, K., Evans, M., Eby, L. T., & Dolan, E. L. (2016). A social capital perspective on the mentoring of undergraduate life science researchers: An empirical study of undergraduate–postgraduate–faculty triads. *CBE—Life Sciences Education*, *15*(2), 1–15. doi: 10.1187/cbe.15-10-0208

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

Altman, D., Machin, D., Bryant, T., & Gardner, M. (Eds.). (2000). *Statistics with confidence* (2nd ed.). Great Britain: BMJ Books.

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

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., … & Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, *13*(1), 29–40. doi: 10.1187/cbe.14-01-0004

Ayella, A., & Beck, M. R. (2018). A course-based undergraduate research experience investigating the consequences of nonconserved mutations in lactate dehydrogenase. *Biochemistry and Molecular Biology Education*, *46*(3), 285–296. doi: 10.1002/bmb.21115

Baker, V. L., Pifer, M. J., Lunsford, L. G., Greer, J., & Ihas, D. (2015). Faculty as mentors in undergraduate research, scholarship, and creative work: Motivating and inhibiting factors. *Mentoring and Tutoring: Partnership in Learning*, *23*(5), 394–410. doi: 10.1080/13611267.2015.1126164

Bakshi, A., Patrick, L. E., & Wischusen, E. W. (2016). A framework for implementing course-based undergraduate research experiences (CUREs) in freshman biology labs. *American Biology Teacher*, *78*(6), 448–455. doi: 10.1525/abt.2016.78.6.448.448

Ballen, C. J., Blum, J. E., Brownell, S., Hebert, S., Hewlett, J., Klein, J. R., … & Cotner, S. (2017). A call to develop course-based undergraduate research experiences (CUREs) for nonmajors courses. *CBE—Life Sciences Education*, *16*(2), 1–7. doi: 10.1187/cbe.16-12-0352

Bathgate, M. E., Aragón, O. R., Cavanagh, A. J., Waterhouse, J. K., Frederick, J., & Graham, M. J. (2019). Perceived supports and evidence-based teaching in college STEM. *International Journal of STEM Education*, *6*(11), 1–14.

Brownell, S. E., Hekmat-Scafe, D. S., Singla, V., Seawell, P. C., Imam, J. F. C., Eddy, S. L., … & Cyert, M. S. (2015). A high-enrollment course-based undergraduate research experience improves student conceptions of scientific thinking and ability to interpret data. *CBE—Life Sciences Education*, *14*(2), ar21. doi: 10.1187/cbe.14-05-0092

Brownell, S. E., & Kloser, M. J. (2015). Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology. *Studies in Higher Education*, *40*(3), 525–544. doi: 10.1080/03075079.2015.1004234

Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. J. (2013). Context matters: Volunteer bias, small sample size, and the value of comparison groups in the assessment of research-based undergraduate introductory biology lab courses. *Journal of Microbiology & Biology Education*, *14*(2), 176–182. doi: 10.1128/jmbe.v14i2.609

Burns, J. C., Okey, J. R., & Wise, K. C. (1985). Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching*, *22*(2), 169–177. doi: 10.1002/tea.3660220208

Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, *26*, 3661–3675.

Choi, K., Lee, H., Shin, N., Kim, S. W., & Krajcik, J. (2011). Re-conceptualization of scientific literacy in South Korea for the 21st century. *Journal of Research in Science Teaching*, *48*(6), 670–697. doi: 10.1002/tea.20424

Cole, M. F., Hickman, M. A., Morran, L., & Beck, C. W. (2021). Assessment of course-based research modules based on faculty research in introductory biology. *Journal of Microbiology & Biology Education*, *22*(2), 1–9. doi: 10.1128/jmbe.00148-21

Connell, G. L., Donovan, D. A., & Chambers, T. G. (2016). Increasing the use of student-centered pedagogies from moderate to high improves student learning and attitudes about biology. *CBE—Life Sciences Education*, *15*(1), 1–15. doi: 10.1187/cbe.15-03-0062

Cooper, K. M., Blattman, J. N., Hendrix, T., & Brownell, S. E. (2019). The impact of broadly relevant novel discoveries on student project ownership in a traditional lab course turned CURE. *CBE—Life Sciences Education*, *18*(4), 1–14. doi: 10.1187/cbe.19-06-0113

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, *14*(1), 1–13. doi: 10.1187/cbe.14-10-0167

Corwin, L. A., Runyon, C. R., Ghanem, E., Sandy, M., Clark, G., Palmer, G. C., … & Dolan, E. L. (2018). Effects of discovery, iteration, and collaboration in laboratory courses on undergraduates' research career intentions fully mediated by student ownership. *CBE—Life Sciences Education*, *17*(2), 1–11. doi: 10.1187/cbe.17-07-0141

Couch, B. A., Brown, T. L., Schelpat, T. J., Graham, M. J., & Knight, J. K. (2015). Scientific teaching: Defining a taxonomy of observable practices. *CBE—Life Sciences Education*, *14*(1), 1–12. doi: 10.1187/cbe.14-01-0002

Cuthbert, D., Arunachalam, D., & Licina, D. (2012). "It feels more important than other classes I have done": An "authentic" undergraduate research experience in sociology. *Studies in Higher Education*, *37*(2), 129–142. doi: 10.1080/03075079.2010.538473

Dirks, C., & Cunningham, M. (2006). Enhancing diversity in science: Is teaching science process skills the answer? *CBE—Life Sciences Education*, *5*(3), 218–226. doi: 10.1187/cbe.05-10-0121

Dolan, E., & Johnson, D. (2009). Toward a holistic view of undergraduate research experiences: An exploratory study of impact on graduate/postdoctoral mentors. *Journal of Science Education and Technology*, *18*(6), 487–500. doi: 10.1007/s10956-009-9165-3

Dolan, E., & Johnson, D. (2010). The Undergraduate–postgraduate–faculty triad: Unique functions and tensions associated with undergraduate research experiences at research universities. *CBE—Life Sciences Education*, *9*(2), 543–553. doi: 10.1187/cbe.10

Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sciences Education*, *17*(1), 1–13. doi: 10.1187/cbe.17-04-0066

Eva, K. W., Cunnington, J. P. W., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in Health Sciences Education*, *9*(3), 211–224. doi: 10.1023/B:AHSE.0000038209.65714.d4

Felege, C. J., & Ralph, S. G. (2019). Evaluating the efficacy of a student-centered active learning environment for undergraduate programs (SCALE-UP) classroom for major and non-major biology students. *Journal of Biological Education*, *53*(1), 98–109. doi: 10.1080/00219266.2018.1447001

Fendos, J. (2018). US experiences with STEM education reform and implications for Asia. *International Journal of Comparative Education and Development*, *20*(1), 51–66. doi: 10.1108/IJCED-10-2017-0026

Fendos, J. (2020a). Anatomy terminology performance is improved by combining jigsaws, retrieval practice, and cumulative quizzing. *Anatomical Sciences Education*, *17*, 1–17. doi: 10.1002/ase.2018

Fendos, J. (2020b). Why a grassroots Summer Institutes model failed: Exploring obstacles to evidence-based teaching awareness in low-awareness, low-support contexts. *Biochemistry and Molecular Biology Education*, *48*, 143–155. doi: 10.1002/bmb.21325

Fendos, J. (2021). Combining jigsaws, rule-based learning, and retrieval practice improves IUPAC nomenclature competence. *Journal of Chemical Education*, *98*(5), 1503–1517. doi: 10.1021/acs.jchemed.0c01235

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*(1), 80–92. doi: 10.1177/160940690600500107

Feyzioglu, B. (2012). Developing a science process skills test for secondary students: Validity and reliability study. *Educational Sciences: Theory & Practice*, *12*(3), 1899–1906.

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

Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, *39*(1–2), 35–49. doi: 10.1016/S0883-0355(03)00072-7

Goodwin, E. C., Anokhin, V., Gray, M. J., Zajic, D. E., Podrabsky, J. E., & Shortlidge, E. E. (2021a). Is this science? Students' experiences of failure make a research-based course feel authentic. *CBE—Life Sciences Education*, *20*(1), 1–15. doi: 10.1187/cbe.20-07-0149

Goodwin, E. C., Cary, J. R., & Shortlidge, E. E. (2021b). Enthusiastic but inconsistent: Graduate teaching assistants' perceptions of their role in the CURE classroom. *CBE—Life Sciences Education*, *20*(4), 1–14. doi: 10.1187/cbe.21-04-0106

Gross, M. M., Wright, M. C., & Anderson, O. S. (2017). Effects of image-based and text-based active learning exercises on student examination performance in a musculoskeletal anatomy course. *Anatomical Sciences Education*, *10*(5), 444–455. doi: 10.1002/ase.1684

Gulacar, O., Sinan, O., Bowman, C. R., & Yildirim, Y. (2015). Exploring the changes in students' understanding of the scientific method using word associations. *Research in Science Education*, *45*(5), 717–726. doi: 10.1007/s11165-014-9443-9

Huitt, T. W., Killins, A., & Brooks, W. S. (2015). Team-based learning in the gross anatomy laboratory improves academic performance and students' attitudes toward teamwork. *Anatomical Sciences Education*, *8*(2), 95–103. doi: 10.1002/ase.1460

Hung, D., Jamaludin, A., & Toh, Y. (2015). Apprenticeship, epistemic learning, and diffusion of innovations in education. *Educational Technology*, *55*(4), 20–26.

Hunter, A., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education*, *91*(1), 36–74. doi: 10.1002/sce

Hurtado, S., Eagan, M. K., Tran, M. C., Newman, C. B., Chang, M. J., & Velasco, P. (2011). "We do science here": Underrepresented students' interactions with faculty in different college contexts. *Journal of Social Issues*, *67*(3), 553–579. doi: 10.1111/j.1540-4560.2011.01714.x

Irby, S. M., Pelaez, N. J., & Anderson, T. R. (2018). How to identify the research abilities that instructors anticipate students will develop in a biochemistry

course-based undergraduate research experience (CURE). *CBE—Life Sciences Education*, *17*(2), 1–14. doi: 10.1187/cbe.17-12-0250

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. doi: 10.1111/j.1365-2929.2004.02012.x

Jones, C. K., & Lerner, A. B. (2019). Implementing a course-based undergraduate research experience to grow the quantity and quality of undergraduate research in an animal science curriculum. *Journal of Animal Science*, *97*(11), 4691–4697. doi: 10.1093/jas/skz319

Keller, G. E. (2002). Using problem-based and active learning in an interdisciplinary science course for non-science majors. *Journal of General Education*, *51*(4), 272–281. doi: 10.1353/jge.2003.0013

Kloser, M. J., Brownell, S. E., Chiariello, N. R., & Fukami, T. (2011). Integrating teaching and research in undergraduate biology laboratory education. *PLoS Biology*, *9*(11), 9–11. doi: 10.1371/journal.pbio.1001174

Kramer, M., Olson, D., & Walker, J. D. (2018). Design and assessment of online, interactive tutorials that teach science process skills. *CBE—Life Sciences Education*, *17*(2), 1–11. doi: 10.1187/cbe.17-06-0109

Limeri, L. B., Asif, M. Z., Bridges, B. H. T., Esparza, D., Tuma, T. T., Sanders, D., … & Dolan, E. L. (2019). "Where's my mentor?!" Characterizing negative mentoring experiences in undergraduate life science research. *CBE—Life Sciences Education*, *18*(4), 1–13. doi: 10.1187/cbe.19-02-0036

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. *Science*, *347*(6222), 628–632. doi: 10.1126/science.1261757

Lopatto, D., Alvarez, C., Barnard, D., Chandrasekaran, C., Chung, H., Du, C., … & Elgin, S. (2008). Genomics Education Partnership. *Science*, *322*, 684–685.

Mangiafico, S. S. (2015). Two-sample Mann–Whitney U Test. In *An R companion for the Handbook of Biological Statistics*. Retrieved May 15, 2019, from http://rcompanion.org/handbook/F_04.html

McGill, T. L., Williams, L. C., Mulford, D. R., Blakey, S. B., Harris, R. J., Kindt, J. T., … & Powell, N. L. (2019). Chemistry unbound: Designing a new four-year undergraduate curriculum. *Journal of Chemical Education*, *96*, 35–46. doi: 10.1021/acs.jchemed.8b00585

Moreira, R. F. (2013). A game for the early and rapid assimilation of organic nomenclature. *Journal of Chemical Education*, *90*(8), 1035–1037. doi: 10.1021/ed300473r

Murren, C. J., Wolyniak, M. J., Rutter, M. T., Bisner, A. M., Callahan, H. S., Strand, A. E., & Corwin, L. A. (2019). Undergraduates phenotyping *Arabidopsis* knockouts in a course-based undergraduate research experience: Exploring plant fitness and vigor using quantitative phenotyping methods. *Journal of Microbiology & Biology Education*, *20*(2), 10. doi: 10.1128/jmbe.v20i2.1650

Nakiboglu, C. (2008). Using word associations for assessing non major science students' knowledge structure before and after general chemistry instruction: The case of atomic structure. *Chemistry Education Research and Practice*, *9*(4), 309–322. doi: 10.1039/b818466f

National Research Council. (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press.

Owens, M. T., Trujillo, G., Seidel, S. B., Harrison, C. D., Farrar, K. M., Benton, H. P., … & Tanner, K. D. (2018). Collectively improving our teaching: Attempting biology department–wide professional development in scientific teaching. *CBE—Life Sciences Education*, *17*(1), 1–17. doi: 10.1187/cbe.17-06-0106

Peteroy-Kelly, M. A., Marcello, M. R., Crispo, E., Buraei, Z., Strahs, D., Isaacson, M., … & Zuzga, D. (2017). Participation in a year-long CURE embedded into major core genetics and cellular and molecular biology laboratory courses results in gains in foundational biological concepts and experimental design skills by novice undergraduate researchers. *Journal of Microbiology & Biology Education*, *18*(1), 9–12. doi: 10.1128/jmbe.v18i1.1226

Pfund, C., Miller, S., Brenner, K., Bruns, P., Chang, A., Ebert-May, D., … & Handelsman, J. (2009). Summer Institute to improve university science teaching. *Science*, *324*(5926), 470–471. doi: 10.1126/science.1170015

President's Council of Advisors on Science and Technology. (2012). *Engage and excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Washington, DC: U.S. Government Office of Science and Technology.

Richardson, J. (2011). The analysis of 2 x 2 contingency tables—Yet again. *Statistics in Medicine*, *30*

Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, *52*(6), 847–867. doi: 10.1002/tea.21221

Rodrigo-Peiris, T., Xiang, L., & Cassone, V. M. (2018). A low-intensity, hybrid design between a "traditional" and a "course-based" research experience yields positive outcomes for science undergraduate freshmen and shows potential for large-scale application. *CBE—Life Sciences Education*, *17*(4), 1–18. doi: 10.1187/cbe.17-11-0248

Rowland, S., Pedwell, R., Lawrie, G., Lovie-Toon, J., & Hung, Y. (2016). Do we need to design course-based undergraduate research experiences for authenticity? *CBE—Life Sciences Education*, *15*(4), ar79. doi: 10.1187/cbe.16-02-0102

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of undergraduate research experiences. *Science*, *316*, 548–549.

Savery, J. R. (2015). Overview of problem-based learning: Definitions and distinctions. In Walker, A., Leary, H., Hmelo-Silver, C., & Ertmer, P., (Eds.), *Essential readings in problem-based learning* (pp. 5–15). West Lafayette, IN: Purdue University Press.

Schiekirka, S., Reinhardt, D., Beibarth, T., Anders, S., Pukrop, T., & Raupach, T. (2013). Estimating learning outcomes from pre-and posttest student self-assessments: A longitudinal study. *Academic Medicine*, *88*(3), 369–375. doi: 10.1097/ACM.0b013e318280a6f6

Sewall, J. M., Oliver, A., Denaro, K., Chase, A. B., Weihe, C., Lay, M., … & Whiteson, K. (2020). Fiber Force: A fiber diet intervention in an advanced course-based undergraduate research experience (CURE) course. *Journal of Microbiology & Biology Education*, *21*(1), 110. doi: 10.1128/jmbe.v21i1.1991

Shaffer, C., Alvarez, C. J., Bednarski, A. E., Dunbar, D., Goodman, A. L., Reinke, C., … & Elgin, S. C. R. (2014). A course-based research experience: How benefits change with increased investment in instructional time. *CBE—Life Sciences Education*, *13*(1), 111–130. doi: 10.1187/cbe-13-08-0152

Spell, R. M., Guinan, J. A., Miller, K. R., & Beck, C. W. (2014). Redefining authentic research experiences in introductory biology laboratories and barriers to their implementation. *CBE—Life Sciences Education*, *13*(1), 102–110. doi: 10.1187/cbe.13-08-0169

Staub, N. L., Poxleitner, M., Braley, A., Smith-Flores, H., Pribbenow, C. M., Jaworski, L., … & Anders, K. R. (2016). Scaling up: Adapting a phage-hunting course to increase participation of first-year students in research. *CBE—Life Sciences Education*, *15*(2), 1–11. doi: 10.1187/cbe.15-10-0211

Stoltzfus, J. R., & Libarkin, J. (2016). Does the room matter? Active learning in traditional and enhanced lecture spaces. *CBE—Life Sciences Education*, *15*(4), 1–10. doi: 10.1187/cbe.16-03-0126

Stuckey, H. (2014). The second step in data analysis: Coding qualitative research data. *Journal of Social Health and Diabetes*, *03*(01), 007–010. doi: 10.4103/2321-0656.140875

Suchman, E. L. (2014). Changing academic culture to improve undergraduate STEM education. *Trends in Microbiology*, *22*(12), 657–659. doi: 10.1016/j.tim.2014.09.006

Thiry, H., & Laursen, S. L. (2011). The role of student-advisor interactions in apprenticing undergraduate researchers into a scientific community of practice. *Journal of Science Education and Technology*, *20*(6), 771–784. doi: 10.1007/s10956-010-9271-2

Thiry, H., Weston, T. J., Laursen, S. L., & Hunter, A. B. (2012). The benefits of multi-year research experiences: Differences in novice and experienced students' reported gains from undergraduate research. *CBE—Life Sciences Education*, *11*(3), 260–272. doi: 10.1187/cbe.11-11-0098

Van Merriënboer, J. J. G., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies. *Medical Education*, *44*(1), 85–93. doi: 10.1111/j.1365-2923.2009.03498.x

Walcott, R. L., Corso, P. S., Rodenbusch, S. E., & Dolan, E. L. (2018). Benefit-cost analysis of undergraduate education programs: An example analysis of the freshman research initiative. *CBE—Life Sciences Education*, *17*(1), 1–8. doi: 10.1187/cbe.17-06-0114

Warfa, A. R. M. (2016). Using cooperative learning to teach chemistry: A meta-analytic review. *Journal of Chemical Education*, *93*(2), 248–255. doi: 10.1021/acs.jchemed.5b00608

Wei, C. A., & Woodin, T. (2011). Undergraduate research experiences in biology: Alternatives to the apprenticeship model. *CBE—Life Sciences Education*, *10*(2), 123–131. doi: 10.1187/cbe.11-03-0028

Ziegler, B., & Montplaisir, L. (2014). Student perceived and determined knowledge of biology concepts in an upper-level biology course. *CBE—Life Sciences Education*, *13*(2), 322–330. doi: 10.1187/cbe.13-09-0175

Zydney, J. M., Warner, Z., & Angelone, L. (2020). Learning through experience: Using design based research to redesign protocols for blended synchronous learning environments. *Computers and Education*, *143*(September 2019), 103678. doi: 10.1016/j.compedu.2019.103678