The Biophysicist

Biophysics Concept Inventory Survey: An Assessment in Biophysical Undergraduate Education

Danielle R. Latham[1], **Joshua D. Alper**[1,2,3,†], **Hugo Sanabria**[1,*]

[1]Department of Physics and Astronomy, Clemson University, Clemson, SC, USA
[2]Department of Biological Sciences, Clemson University, Clemson, SC, USA
[3]Eukaryotic Pathogen Innovations Center, Clemson University, Clemson, SC, USA

ABSTRACT Since the Force Concept Inventory in 1992, many concept inventories have been developed to cover classical scientific fields. However, there is a lack of concept inventories for interdisciplinary fields, such as biophysics. We introduce a Biophysics Concept Inventory Survey (BCIS), a 20-question, multiple-choice survey to measure student gains in biophysical concepts. The BCIS contains 5 question classifications: remember, understand, apply, analyze, and create, as well as question concepts divided into primarily physics or primarily biology questions. We administered the BCIS to 3 cohorts of students over 4 years. Each cohort participated in a 10-week summer Research Experience for Undergraduates (REU) in biophysics. We compared the presurvey (before REU) and postsurvey (after REU) scores to determine the fraction of the maximum possible gain or loss realized. Our analysis of the results suggests that the BCIS shows no biases based on sex or ethnicity. Further, we used the BCIS to show that 69% of the REU participants showed gains in biophysics concepts, with most of the total participant mean of gain occurring at higher levels of Bloom's taxonomy: create and analyze. Overall, participants obtain slightly higher scores in physics (8% increase) than biology (5% increase) when comparing the pre- and postscores. The coronavirus disease 2019 pandemic allows a splitting of prepandemic and postpandemic cohorts, with the postpandemic cohort showing significantly larger gains than the prepandemic students. These results show that the BCIS, with question classifications and concepts, probes the students' ability to apply knowledge to various biophysical science topics without underlying biases and enables instructors to obtain answers to important questions about the effectiveness of the educational programs. The BCIS fills a gap for interdisciplinary concept inventories.

KEY WORDS concept inventory; biophysics education, survey; testing metrics; gains; effect size; Bloom taxonomy; experiential learning; research-driven learning

I. INTRODUCTION

Concept inventories exist for traditional fields in science (1–3), technology (4, 5), engineering (6, 7), and mathematics (8, 9), but there remains a gap for interdisciplinary fields such as biophysics. Concept inventories arise from the need to have metrics to determine the depth of common student misunderstandings in sciences. Multiple-choice questions allow instructors a quick, easy-to-grade method of probing classes for complex topics. Several studies have already shown the benefits and logic of concept inventories (10) and how they can be best applied (11–14).

We developed a Biophysics Concept Inventory Survey (BCIS) to probe student learning across disciplines by generating 20 multiple-choice questions, which take an interdisciplinary approach to physics and biology. The BCIS contains 5 question classifications based on Bloom's taxonomy:

"†" current address: Magnet Biomedicine, Boston, MA, USA

"*" corresponding author

Received: 9 August 2023
Accepted: 20 May 2024
Published: 18 July 2024

© 2024 Biophysical Society.

https://thebiophysicist.kglmeridian.com | 2025-01-18

remember, understand, apply, analyze, and create (13, 14). The question classification allows the probing of students' ability to apply biophysical concepts to various problems. Moreover, we classified the questions as either prominently probing physical or biological concepts. Results inform instructors in which concepts students struggle the most. We tested the BCIS for underlying biases by using gathered demographic information, including sex and ethnicity of the Research Experience for Undergraduates (REU) participants.

II. SCIENTIFIC AND PEDAGOGIC BACKGROUND

Students have many incorrect ideas and misconceptions regarding science (15–17). Written exams and student interviews help determine these misconceptions, but they are long and take time to perform and analyze. In 1985, Halloun and Hestenes developed a multiple-choice concept inventory regarding the physics of motion to quickly determine student misconceptions (18). Questions for this survey were multiple-choice, with 1 correct answer and several incorrect answers designed to distract. These distractor answers were designed from common misconceptions based on common student answers in written essays and student interviews (19). Shortly after the motion concept inventory, the Force Concept Inventory was developed (20). The Force Concept Inventory showed that students could recite Newton's third law but not apply it correctly. These early concept inventories led to an overturn of science education (10, 12).

Since the first release of concept inventories, particularly the Force Concept Inventory, there have been several studies showing the benefits and logic of concept inventories and how they can be best applied (11, 20–22). Multiple-choice questions allow administrators a quick, easy-to-grade means of probing student learning in complex topics. Concept inventories serve as a valuable tool for assessing the level of student comprehension and misconceptions in the field of sciences (17, 23, 24).

Administrating concept inventories several times throughout a course allows instructors to determine student education progress during instruction. Typically, this change involves students doing better on the concept inventory after instruction showing an increase in score. The change in score gives a measure of how much information students gain after instruction. Often gain is the metric used to determine student advancement in a course. Although there are ongoing discussions regarding the best way to calculate gain (25, 26), gain is typically on a scale between 0 and 1, with a traditional, semester-long lecture course giving an average gain of $\sim$0.25 (12).

The work of Halloun and Hestenes helped guide the creation of future concept inventories, giving way to numerous concept inventories in multiple disciplines, including physics (20, 21, 27, 28), chemistry (29, 30), and biology (1, 2, 31–34). However, these concept inventories are very specific, often covering a single topic within a single discipline such as kinematics (27) or electrostatics (28) from physics and natural selection (2) from biology. There are several concept inventories for traditional fields, but there remains a lack of tools for measuring student learning and understanding in interdisciplinary fields, such as biophysics. We developed a BCIS to address this need.

We developed the BCIS to assess student understanding across disciplines by generating 20 multiple-choice questions that take an interdisciplinary approach to physics and biology. We wrote questions to be classified as primarily physics-based or primarily biology-based topics to inform instructors about topics that cause students to struggle. Physics questions are typical physics concepts, including diffusion, kinetics, force and energy, density, pressure, mechanics, electrostatics, and optics, applied to a biological system, such as switching a walking person in a kinetics question to a cargo vesicle moving along a microtubule. Biology questions put core biological concepts in the front, including molecular biology, genetics, and biochemistry, with less emphasis on the physical properties of biomolecules. As guidance, we modified some questions from previously existing concept inventories. For example, 1 question comes from the Force Concept Inventory (20), where the original

https://thebiophysicist.kglmeridian.com | 2025-01-18

question involved the forces between charged spheres. Our modified version creates the situation as proteins are embedded in a cellular membrane. Biology-based questions came from general biology concepts being more definition based and mechanism driven.

The additional design of the BCIS included considering Bloom's taxonomy of human cognition. We group each BCIS question into 1 of 5 classifications: remember, understand, apply, analyze, and create (13). These classifications enable instructors to probe the student's ability to apply biophysical concepts at various cognition levels. It is not enough to repeat previous facts, but students should be able to use knowledge to further research and assist with troubleshooting. We want students to form problem solving and logic skills. Addressing the questionnaire as a survey helps students answer honestly and address test anxiety (35, 36).

We calculated each participant's gain or loss of knowledge and then averaged the gains and losses together for an average of gains. We tested the BCIS for biases against sex and ethnicity. There were no significant differences between sex and ethnicity. Our study was interrupted by the coronavirus disease 2019 (COVID-19) pandemic. This interruption allowed us a unique opportunity to demonstrate how the BCIS distinguished between pre- and postpandemic cohorts. Our results, for the pilot REU group, imply that the BCIS can be used to determine the change in student understanding and application over time by using multiple-choice questions for quick and easy grading. Thus, the BCIS fulfills a need for interdisciplinary evaluations across biophysics courses. This work, classified as exempt under Category 2 in accordance with the Code of Federal Regulations (CFR) 45 CFR 46.104(d) (37), was carried out in accordance with the standards established by the Clemson University Institutional Review Board (2018270).

III. METHODS AND STATISTICAL TESTS

A. The BCIS

The BCIS consists of 20 multiple-choice questions with a single correct answer. Example questions can be seen in the Supplemental Material. We used the Force Concept Inventory (20) as an example. For questions classified as primarily physics, we used applications of physics concepts to biological systems. For example, instead of a charged particle, we used a charged DNA. Simple explanations were changed to have a biological context. For questions classified as primarily biology, we asked semiquantitative questions focused on molecular biology, genetics, and biochemistry.

Instructor access to the BCIS can be requested by filling out a Google form with proof of the instructors' role (38).

B. The REU sample group

As a pilot test, we administered the BCIS to 32 students from 3 cohorts of undergraduate researchers who participated in the Clemson University REU site ("Nature's Machinery through the Prism of Physics, Biology, Chemistry and Engineering") funded by the National Science Foundation. The REU committee, consisting of the primary investigators of the REU site, and a faculty mentor screened the applications to satisfy the programmatic goals of equal participation from participants with backgrounds in the biological and the physical sciences. For each cohort, the REU committee balanced participation from those of underrepresented minority (URM) status, on the basis of sex, and from nonresearch-intensive institutions. Final assignment to the project was equally weighted by the participant's interest and a final interview with the potential mentor. Recruitment was nationally, but with emphasis from the southeast. The participants came from 17 states, from private and public institutions of higher education, ranging from primarily undergraduate institutions to doctoral universities with very high research activity according to the Carnegie Classification (39). As part of the application, we gathered demographic information on the participants, such as sex and ethnicity.

The first week of the REU program, undergraduate researchers participated in a "Biophysics

https://thebiophysicist.kglmeridian.com | 2025-01-18

Bootcamp." During bootcamp, participants participated in approximately 13 h of traditional lectures and 17 h of laboratory work, including introduction to research lectures. Participants spent this first bootcamp week becoming familiar with Clemson University's campus, socializing with each other, and learning essential research and basic laboratory skills, such as how to keep a laboratory notebook and research safety, among other required introductions before entering a laboratory setting. In addition, each cohort received training in basic experimental and computational tools following a designed theme. For example, in 2021, participants determined the size of green fluorescent protein by various means, including fluorescent correlation spectroscopy, size exclusion chromatography, computational simulations using Visual Molecular Dynamics (40), and quantitative analysis of sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels. After the bootcamp, participants wrote a report formatted as a biophysical journal article. This training helped participants understand experimental validation through many means and determine the differences (pros and cons) of different experimental designs.

For the remaining 9 weeks of the REU program, participants worked on collaborative, interdisciplinary research projects in pairs, but with individual and unique project objectives, where 1 undergraduate researcher had an experimental focus, while the other had a computational aspect of the same problem; or 1 undergraduate researcher was in a physics laboratory, and the other was doing the more biological aspects of the project. This approach allowed participants to build collaboration skills, while gaining exposure to both experimental and computational approaches to research.

To supplement the experience and aid in building professional development skills (41, 42), REU participants had weekly meetings with cohorts where they presented research updates, including project design, background, and scientific importance. Participants also met weekly for a journal club and took turns presenting recently published research articles relating to the project to encourage staying up-to-date on relevant research for the topic and practicing critical reading of the literature. There were also weekly professional seminars given by experts at the university covering topics such as scientific writing, networking, and conflict resolution. At the end of the summer, undergraduate researchers participated in Clemson University's undergraduate research symposium.

The goals of the REU were to (a) encourage and enable participants to pursue interdisciplinary research careers, (b) provide participants with important and feasible projects done and mentored collaboratively by biological and physical scientists, (c) train participants to communicate science clearly, and (d) provide career development advice, research skills, and mentorship. As such, REU participants did not have any traditional classroom instruction regarding the topics covered by the BCIS, and participants were not quizzed or given traditional homework, such as problem sets. The BCIS was developed separately from the REU curriculum. Participants drove learning by finding and reading scientific literature, asking questions of those around them, and problem solving research projects. Thus, this sampling is biased toward undergraduate researchers who participated in an interactive, experiential learning approach (22, 43), instead of students who participated in a traditional, semester-long lecture course.

C. Administration of the survey

Participants took the BCIS upon arrival (presurvey) to the REU site and upon departure (postsurvey). The question order remained the same for the presurvey and postsurvey to ensure the order of the question played no part in answer changes between pre- and postsurveys. Access to the survey required a password and Respondus LockDown Browser (Version 1.0.5; Redmond, WA) to ensure the survey was given to all participants simultaneously with no outside resources. Participants had 35 min to answer the 20 questions.

https://thebiophysicist.kglmeridian.com | 2025-01-18

D. Matched data

We used matched data (44) for all analyses, allowing the consideration of participant demographics. Therefore, participant data calculations are completed for each individual and then pooled via demographics to form statistical groups.

E. Fraction of maximum possible gain realized

For each participant, we calculated the pre- and postscores from the pre- and postsurvey, respectively. Each question was weighted the same with typical grading procedures to determine the score; the number of correctly answered questions was divided by the total number of questions to give a percentage answered correctly.

Also, Eq. 1 shows how we compared the pre- and postscores for each participant to obtain a gain, no change, loss (GNL) value. We calculated the fraction of the maximum possible gain realized (gain; 12) for participants who scored higher on the postsurvey than the presurvey. For participants who scored lower on the postsurvey than the presurvey, we calculated the maximum possible loss forfeited (loss). Although the concept of loss has been deliberated before (25, 26), our loss calculation method is normalized regarding the percentage of questions answered incorrectly compared with what was initially known. The participant is assigned 0 when the pre- and postscores are identical, signaling no change. There were no participants who obtained a perfect score (100%) on the pre- or postsurvey. Therefore, the GNLs are calculated as follows:

$$\begin{gathered}
\text{postscore} > \text{prescore:} \\
\text{gain} = \frac{\text{postscore} - \text{prescore}}{100\% - \text{prescore}} \\
\text{no change} = 0, \text{postscore} = \text{prescore} \\
\text{postscore} < \text{prescore:} \\
\text{loss} = \frac{\text{postscore} - \text{prescore}}{\text{prescore}}
\end{gathered} \tag{1}$$

With a mean of GNL (gain, G; no change, N; loss, L) that is the weighted average of the 3 possible scores as

$$\begin{aligned}
\langle \text{GNL} \rangle &= \frac{n_G}{n_T}\left(\frac{1}{n_G}\sum_{n_G} G\right) + \frac{n_N}{n_T}\left(\frac{1}{n_N}\sum_{n_N} N\right) \\
&\quad + \frac{n_L}{n_T}\left(\frac{1}{n_L}\sum_{n_L} L\right) \\
&= \frac{1}{n_T}\left(\sum_{n_G} G + \sum_{n_N} N + \sum_{n_L} L\right)
\end{aligned} \tag{2}$$

The mean GNL method creates a scale from −1 (total loss) to 1 (total gain), where negative numbers represent loss and positive numbers represent gain. This method assists in averaging statistics and further data analysis.

F. The *P* values and effect size

Participants were deidentified and grouped into different demographic groups: sex, URM status, and college major, as were self-reported by the participants. We calculated and considered the Cohen effect size (d; Eq. 3; 45–47) to compare between groups. The effect size shows the size of the shift between the pre- and postscores. We opted for the Cohen effect size because it provides a good measure for smaller sampling sizes, and we have a total sample size of 32. The effect size is calculated by Eq. 3, where pooledSTD is the pooled standard deviation of all the pre- and postsurvey scores.

$$\text{effect size} = d = \frac{\langle \text{postscore} \rangle - \langle \text{prescore} \rangle}{\text{pooledSTD}} \tag{3}$$

where the brackets ⟨ ⟩ represent the mean. In this manner, an effect size of 0.2 is a small shift, 0.5 is a medium shift, and 0.8 is a large shift (45, 48).

Further, each demographic grouping was compared by using Student *t* test. For each *t* test, we used normal quantile–quantile plots to ensure the sampling data distribution was close to normal. With such a small sampling size, a *P* value may not be efficient for determining the differences between subgroups (49), but a combination of *P* values and effect size allows a complete comparison between various subgroups for this

https://thebiophysicist.kglmeridian.com | 2025-01-18

study (50). We considered $P < 0.10$ to be statistically significant.

G. Question subject percentages

To determine the effect size regarding subject matter, questions of similar subjects were grouped together. The total mean and standard deviation for the pre- and postresponses were determined for each group. This allowed the pooled standard deviation and effect size for each grouping to be determined.

IV. RESULTS AND DISCUSSION

A. The BCIS shows medium gains from REU participants

Overall, the average BCIS scores increased by 7% from the prescore (49.4% ± 14.2%) to the postscore (56.4% ± 13.1%). With a prescore of 50% (the mean prescore is not statistically different from 50%, $P = 0.8$), the BCIS is easy enough for undergraduate students to feel confident, while leaving enough room for students to achieve gain.

We found that the 32 REU participants had a mean GNL of 0.13 ± 0.18 and an effect size of 0.51 with 3 groups: gain with 22 participants, loss with 6 participants, and no change with 4 participants (Fig 1). Gain participants have a large effect size of 0.98, with a mean of gains 0.23 ± 0.11 ($n = 22$). Loss participants have a medium effect size of −0.40, with a mean of loss −0.15 ± 0.07 ($n = 6$).

The increase in gain and effect size may be attributed to interactive experimental learning and may not reflect a traditional lecture course (51, 52). Many previous studies discard the students with losses (53). Here, we decided to divide the gains and losses but show both groups (26). The 69% of participants benefited from the REU, as assessed by the BCIS with overall positive gains.

B. The BCIS can identify students' weak and strong subjects

We analyzed the BCIS results by question subject. Although the questions are interdisciplinary, we classified each question as a principal biology subject (6 questions) or a physics subject

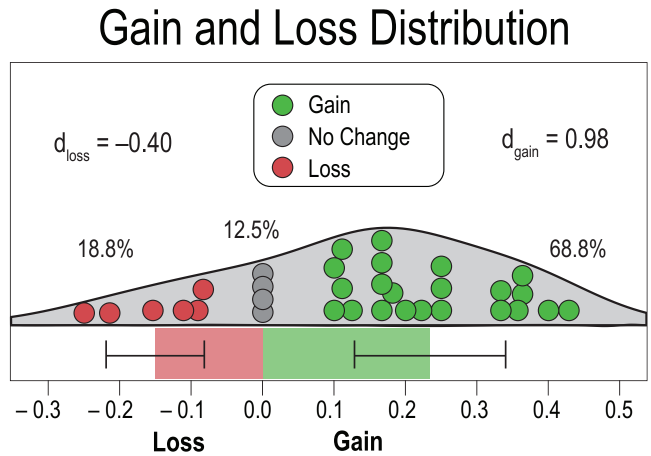

Fig 1. Overall gains and losses. The distribution of the 32 REU students' gain or loss values shown in the horizontal violin chart, where each point represents a student's GNL score (green represents gain, gray represents no change, and red represents loss). The percentage and effect size (*d*) for each group can be found above the respective group. The horizontal bar chart below each group shows the mean with error bars, representing the standard deviation for gain and loss groups.

(14 questions). Further, the questions address specific topics, including kinetics, mechanics, force and energy, electrostatics, density, pressure, diffusion, and optics for physics and molecular biology, genetics, and biochemistry for biology. We compared the pre- and postsurvey responses for each student to identify the topics that individual participants either better understand or continue to struggle with after instruction (Fig 2).

We found that the participants came in to the REU program already understanding biology subjects better than physics subjects, with 69% of answers correct for biology subject questions on the presurvey compared with only 41% for physics subjects. Participants had a slightly larger effect size for physics ($d = 0.16$) compared with biology ($d = 0.12$), but both show small shifts. A closer look showed participants shifted more on certain subjects than others. Within the physics group of questions, we found participants showed larger shifts in introductory physics concepts, such as density ($d = 0.58$), force and energy ($d = 0.45$), and kinetics ($d = 0.31$), with small negative shifts, denoting losses, in more advanced physics concepts, such as electrostatics ($d = -0.11$), and pressure ($d = -0.03$). The small losses may be attributed to guessing, due to the nature of

https://thebiophysicist.kglmeridian.com | 2025-01-18

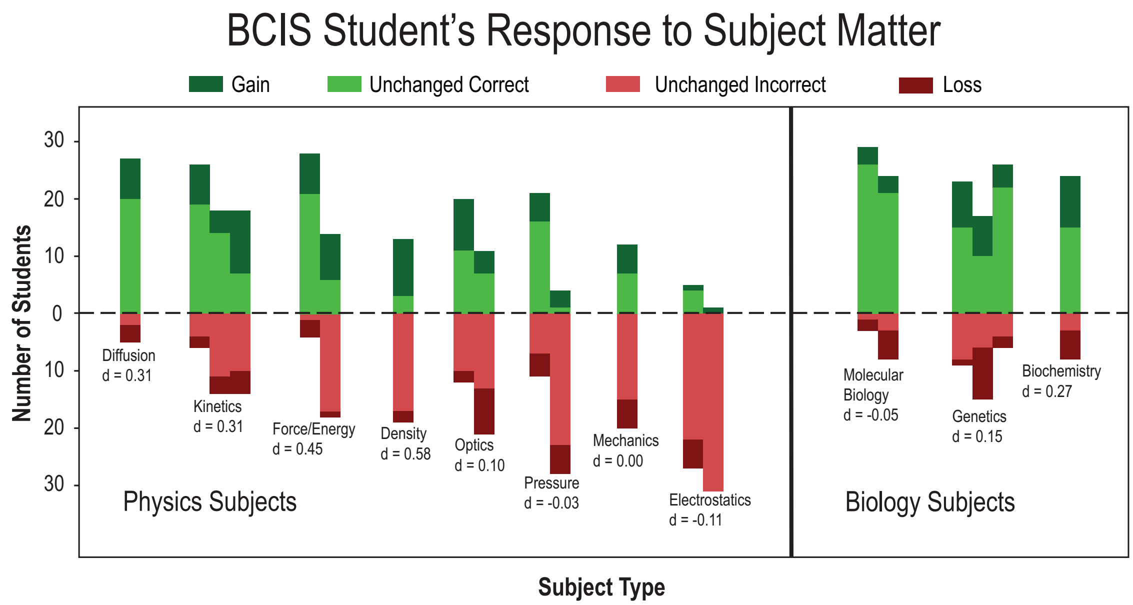

Fig 2. The BCIS contains various question categories that probe a breadth of physics and biology subjects. Bar graphs represent student gain or loss for each question, with labels indicating the main subject the question is covering. Students in dark green are marked as gain, answering correctly only on the postsurvey. Students in light green are marked as unchanged correct, answering correctly on the pre- and postsurvey. Students in light red are marked as unchanged incorrect, answering incorrectly on both the pre- and postsurveys. Students in dark red are marked as loss, answering correctly on the presurvey but incorrectly on the postsurvey.

multiple-choice testing (54). We observe smaller changes regarding biology subjects. Biochemistry ($d = 0.27$) showed the largest total change, with molecular biology ($d = -0.05$) showing a slight negative shift.

C. The BCIS uses question classifications to assess participants' understanding

Typical assessments only tend to probe student knowledge, the ability to repeat information previously given. It is crucial to ensure that students can apply this knowledge. Thus, we designed the BCIS with questions across multiple levels of Bloom's taxonomy of educational objectives (13, 14; Table 1).

We found that the REU students showed nearly 0 mean of GNL at the lower levels of Bloom's taxonomy: remember (0.06 $\pm$ 0.37); understand (0.06 $\pm$ 0.45); and apply (0.05 $\pm$ 0.43). However, we found considerable gains in create (0.22 $\pm$ 0.39) and analyze (0.23 $\pm$ 0.50; Fig 3). We attribute these results to the active learning approach of the REU. The experience increased the participant's ability to apply knowledge, particularly regarding creating new connections (create) and understanding how system parts fit together (analyze). However, without required reading, traditional problem sets, or classroom-based lectures, participant baseline remember showed no gain.

D. This BCIS is nonbiased for sex and URM status but shows a preference for college major

We pooled all REU cohorts by demographic information to test the BCIS for biases (55, 56).

Table 1. Bloom's taxonomy: BCIS question classification details.

Classification	Number of questions	Description of question classification
Remember	7	Recognizing and recalling information
Understand	3	Interpreting, explaining, summarizing
Apply	3	Applying rules, methods, or principles to new situations
Analyze	3	Classifying and understanding components parts within a system
Create	4	Creating new connections and combining ideas
Evaluate	0	Addressing controversies, forming opinions

https://thebiophysicist.kglmeridian.com | 2025-01-18

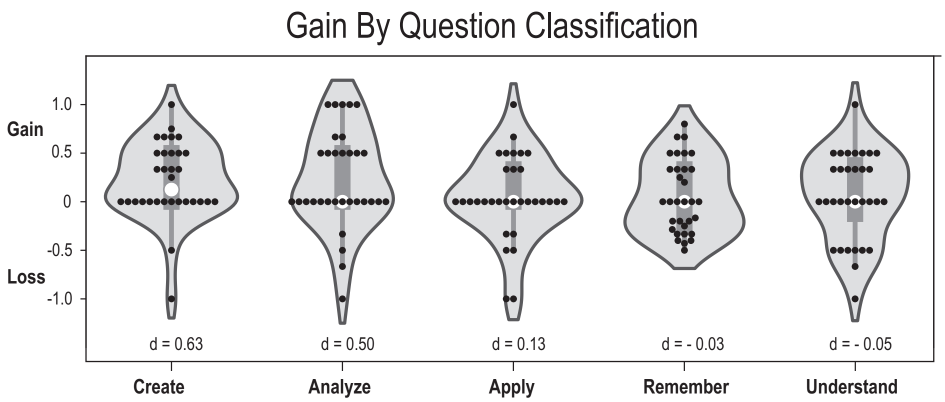

Fig 3. The BCIS contains various question categories that probe different levels of understanding. Violin plots show student gain for each question category. Each black dot represents a student. The box and whisker plots inside the violin plots show the quartiles, with the white dot representing the median GNL. Effect size (*d*) is shown below the plots.

We found no statistically significant differences in gain or loss corresponding to the participants' sex (2-tailed *t* test, $P = 0.90$) or URM status (2-tailed t test, $P = 0.62$). The effect size for sex was 0.41 for males and 0.61 for females, indicating that both groups showed medium gains from the REU (Fig 4A). The effect size for URM status was 0.67 for URM participants and 0.45 for non-URM participants, indicating that both groups also showed medium gains from the REU (Fig 4B). Also, these data support the conclusion that the BCIS is not inherently biased based on gender or ethnicity.

Further, we pooled the participants into respective college majors: physical sciences (for participants majoring in physical sciences or engineering) or biological sciences (for participants majoring in any of the life sciences). Approximately two-thirds

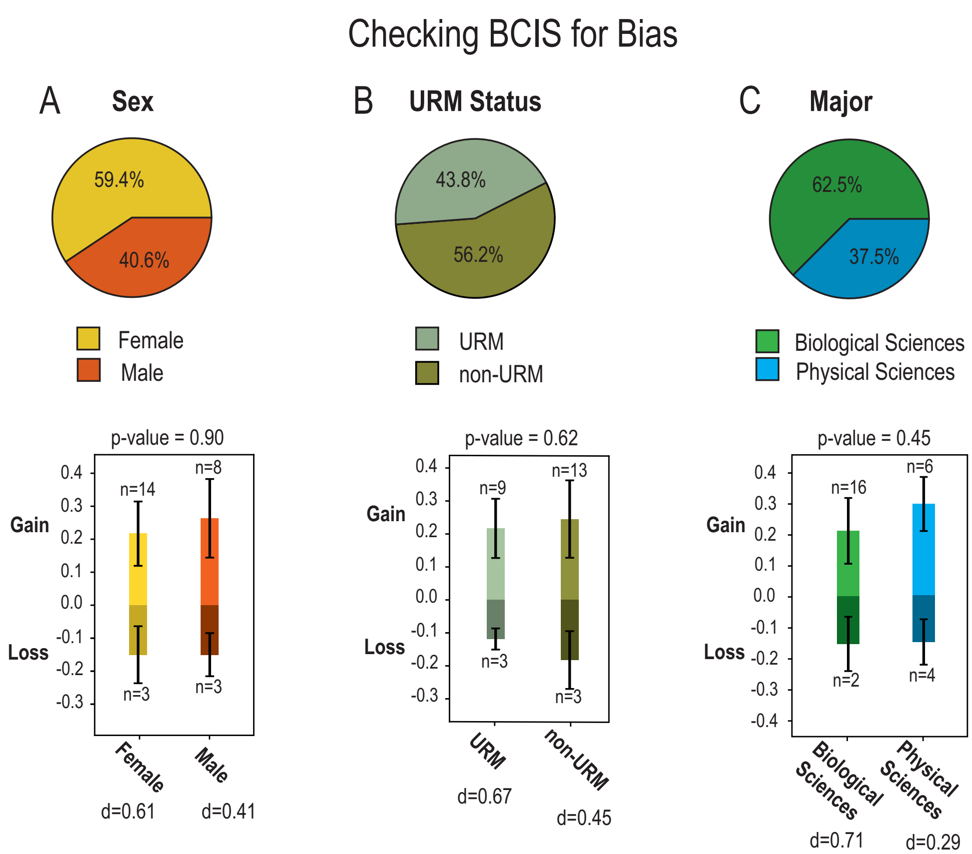

Fig 4. The BCIS shows no bias for sex or URM. Pie charts show the distribution of student demographics based on (A) sex, (B) URM status, and (C) college major. Bar charts show the mean of the gain (positive) and loss (negative), with the error bars denoting the standard deviation for each distribution. The effect size (*d*) is shown below for each demographic. Sample size for each group is denoted by "*n* =" near the error bar for that group. The 4 students with no change are not shown in the bar charts.

https://thebiophysicist.kglmeridian.com | 2025-01-18

of participants had a biological sciences undergraduate major (Fig 4C). We found no statistically significant differences in gain or loss corresponding to the participant's major (2-tailed t test, $P = 0.45$).

However, the effect size is 0.71 for biological sciences, showing large growth, while physical sciences have an effect size of only 0.29, showing small growth that implies that biological science majors experienced bigger gains than physical science majors during the REU. This could be due to biological sciences lacking more physics knowledge than physical sciences lacking biology knowledge at the start of the program.

E. Example: COVID-19 impact on participant gains

We first administered the BCIS to REU participants in 2019; however, a worldwide pandemic interrupted and altered the study's course, as safety concerns postponed the 2020 REU. We offered deferment to those participants we had accepted to the 2020 REU. Thus, the 2021 REU cohort consisted of a mix of participants who were accepted prepandemic (2020; $n = 7$) and postpandemic (2021; $n = 8$). Our analysis of survey data and conversations with participants revealed that the participants who had applied postpandemic (in 2021 and 2022) lacked traditional laboratory courses that would have accompanied the introductory science courses at the home institutes, while those who had applied prepandemic (in 2019 and 2020) had those lab courses. This distinction led us to divide the participant cohorts into prepandemic and postpandemic groups.

We found that the prepandemic group had a mean GNL of 0.07 ± 0.18, with an effect size of 0.35 ($n = 14$), and the postpandemic group had a mean GNL of 0.18 ± 0.18, with an effect size of 0.69 ($n = 18$). A comparison between the 2 groups showed they were significantly different (2-tailed t test, $P = 0.09$). These differences are explained by both larger gains (Fig 5A) and a greater fraction of students showing gain (Fig 5B) in the postpandemic group.

A more detailed inspection of this data using the question classifications (Fig 5C) shows similar, medium effect sizes, indicating gain for both pre- and postpandemic cohorts at the higher levels of Bloom's taxonomy: create and analyze. However, there are significant differences at the lower levels of Bloom's taxonomy, with the prepandemic group showing negative effect sizes in understand and remember, implying a loss. In contrast, the postpandemic group shows small, positive shifts in these classifications.

Together, these results show a distinction between prepandemic and postpandemic cohorts, including a 21% increase in the number of participants who exhibited gain postpandemic, larger effect sizes for questions classified lower on Bloom's taxonomy (understand, remember, apply) for postpandemic participants, and an overall 0.34 increase in effect size and 0.11 increase in gains for postpandemic compared with prepandemic. These results imply educational disruption has interfered with student education, but hands-on, active learning approaches, such as summer REU experiential learning programs, may aid in recovery. They suggest that immersive lab experience benefits students, with the exposure helping return students to a better, prepandemic learning state.

V. CONCLUSION

The current concept inventories are lacking for interdisciplinary fields. To fill this gap, we created the BCIS. We administered the BCIS to 32 REU participants as a pilot group. By having different question classifications and subject material, we could better understand participants' weak points, including second semester physics topics, such as electrostatics, pressure, and optics, as well as applied biological subjects, such as molecular biology. Also, the BCIS results suggest that experiential learning through an REU leads to a higher mean of GNL at the higher end of Bloom's taxonomy (create and analyze) than at the lower end (remember and understand). Applying the BCIS to traditional lecture courses would be interesting, because we anticipate larger gains at the lower end of Bloom's taxonomy. For traditional semester-long ($\sim$16 weeks) courses, it is likely best to apply the BCIS 3 times:

https://thebiophysicist.kglmeridian.com | 2025-01-18

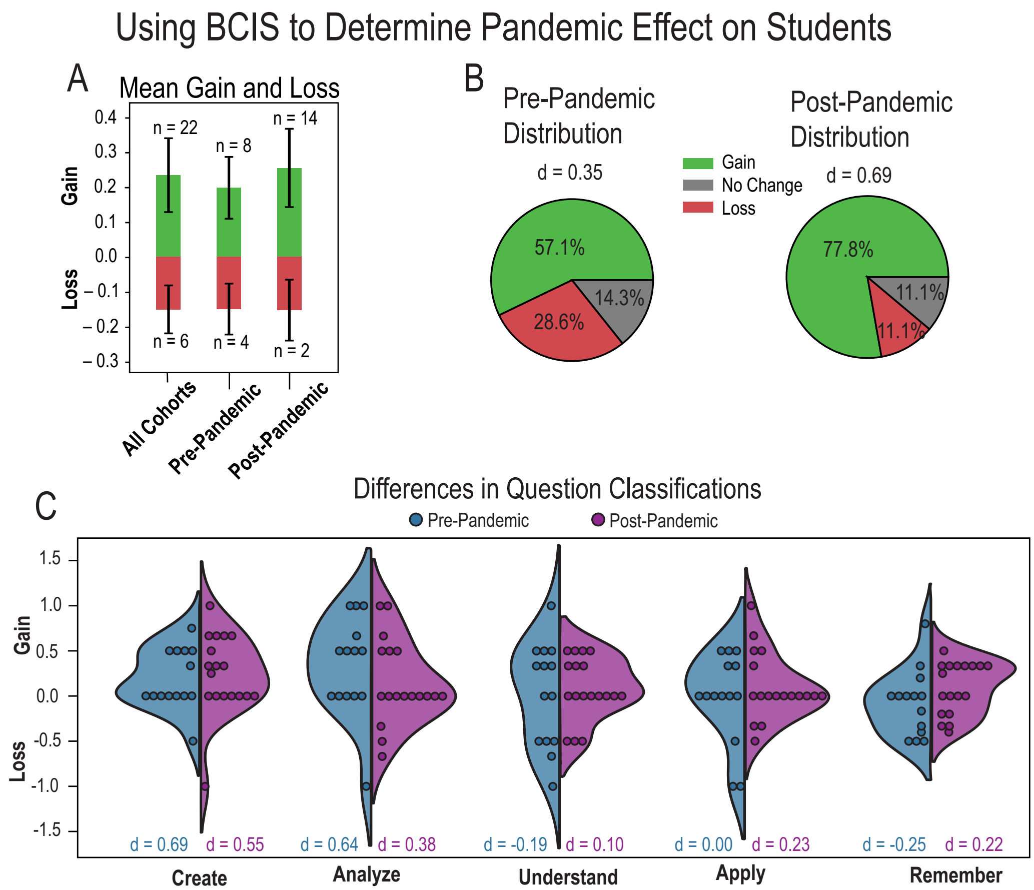

Fig 5. Applying the BCIS before and after the COVID-19 global pandemic suggests an increased benefit of experiential learning opportunities for postpandemic REU students. (A) Bar graph shows mean gain and loss for the pre- and postpandemic cohorts, where the error bars are the standard deviation. (B) Pie charts show fractions of the pre- and postpandemic cohorts with gain and loss. The 4 students with no change are not shown in the bar charts. (C) In the violin plots, each student is represented by a point. The left side (blue) of each violin plot represents the prepandemic cohorts, while the right side (violet) of each violin plot represents the postpandemic cohorts. The effect size (*d*) under each corresponding distribution shows the differences in the pre-and post-BCIS responses for each distribution.

at the start of the semester, halfway through the semester, and at the end of the semester (57). The classification of questions by subject and Bloom's taxonomy level allows instructors to determine what students are struggling with and adjust course direction at a midway point of a course.

This study suggests that surveys such as the BCIS are useful tools to evaluate student gains in interdisciplinary courses and active learning experiences. However, our results are limited to a small number of REU participants. Therefore, we must administer it to more students for a larger sample size. After the BCIS is robustly tested on a larger sample size, many potentials open up, such as (a) building a database of questions, (b) probing class progression at a midpoint, and (c) checking student's previous understanding of physics and biology with topic-specific concept inventories. Instructors who want to apply this to a course or research program can do so by contacting the authors of this study and filling out a Google form with proof of an instructor's role.

In conclusion, the BCIS starts to fill the need for an interdisciplinary method of evaluating student progress in biophysics courses. It is unbiased in measuring interdisciplinary biology and physics understanding. It covers various subjects in physics and biology, allowing understanding of students' weak points. Question classifications based on Bloom's taxonomy grant the ability to understand students' level of knowledge and the ability to apply that knowledge. In our pilot study, we found apparent differences in performance on the BCIS between prepandemic

https://thebiophysicist.kglmeridian.com | 2025-01-18

and postpandemic REU undergraduate researchers. In the future, we will expand the BCIS by adding an extensive data bank of questions, enabling instructors to customize the balancing of the BCIS by classification, subject matter, and question type. A data bank would allow instructors to build a specialized concept inventory for class, covering topics that could be more relevant to specific needs.

VI. LIMITATIONS

This study aimed to introduce the BCIS. The sample selection from REU students results in a small sample size and bias toward students who are self-driven to learn. Further work needs to be completed to ensure the BCIS questions probe the expected concept and are interpreted correctly and validated among a larger pool of participants. The current work presented in this article sets a baseline for interdisciplinary concept inventories and does not include construct validity, content validity, or face validity (58).

SUPPLEMENTAL MATERIAL

Supplemental sample BCIS questions are available at: https://doi.org/10.35459/tbp.2023.000256.S1.

AUTHOR CONTRIBUTIONS

DRL wrote the manuscript, with revision by JDA and HS. DRL did the data analysis. JDA and HS supervised the work and provided guidance to DRL. JDA and HS designed the BCIS. HS lead the administration of the BCIS to the REU participants. JDA and HS are directors of the REU site "Nature's Machinery through the Prism of Physics, Biology, Chemistry and Engineering" at Clemson University.

ACKNOWLEDGMENTS

We thank Celeste Hackett for help with organization of the Research Experience for Undergraduates (REU) site: "Nature's Machinery through the Prism of Physics, Biology, Chemistry and Engineering" funded by the National Science Foundation (NSF). Special thanks to the bootcamp teaching assistances of all cohorts for leading and modifying protocols to match the students' needs: Zhenzhen Zhang, George Hamilton, Katherine Wentworth, Rajen Goutam, Subash Godar, Narendar Kolimi, Ashok Pabbathi, and Vincent Clanzy III. A big thanks to all the primary investigators and guest lectures who mentored the REU participants through the summer at Clemson University. This research was supported by the NSF (award DBI-1757658) to JDA and HS. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. The funders had no role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, and in the preparation, review, or approval of the manuscript. The authors declare no competing interests.

REFERENCES

1. Smith, M. K., W. B. Wood, and J. K. Knight. 2008. The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci Educ* 7:422–430.
2. Anderson, D. L., K. M. Fisher, and G. J. Norman. 2002. Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach* 39:952–978.
3. Evans, D. L., G. L. Gray, S. Krause, J. Martin, C. Midkiff, B. M. Notaros, M. Pavelich, D. Rancour, T. Reed-Rhoads, P. Steif, R. Streveler, and K. Wage. 2003. Progress on concept inventory assessment tools. In Proceedings of the 33rd Annual Frontiers in Education. Westminster, CO, USA. Institute of Electrical and Electronics Engineers. New York, NY, USA. p. T4G-1.
4. Caceffo, R., S. Wolfman, K. S. Booth, and R. Azevedo. 2016. Developing a computer science concept inventory for introductory programming. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education. Memphis, TN, USA. 17 February 2016. Association for Computing Machinery. New York, NY, USA. pp. 364–369.
5. Almstrum, V. L., P. B. Henderson, V. Harvey, C. Heeren, W. Marion, C. Riedesel, L.-K. Soh, and A. E. Tew. 2006. Concept inventories in computer science for the topic discrete mathematics. In Working Group Reports on ITiCSE on Innovation and Technology in Computer Science Education (ITiCSE-WGR '06). Vol. 38, issue 4. pp. 132–145.
6. Krause, S., J. C. Decker, and R. Griffin. 2003. Using a materials concept inventory to assess conceptual gain in introductory materials engineering courses. In Proceedings of the 33rd Annual Frontiers in Education. Westminster, CO, USA. 4 November 2003. Institute of Electrical and Electronics Engineers. New York, NY, USA. p. T3D-7.
7. Prince, M., M. Vigeant, and K. Nottis. 2012. Development of the heat and energy concept inventory: preliminary results on the prevalence and persistence of engineering students' misconceptions. *J Eng Educ* 101:412–438.
8. Steif, P. S., and J. A. Dantzler. 2005. A statics concept inventory: development and psychometric analysis. *J Eng Educ* 94:363–371.
9. Gleason, J., S. Bagley, M. Thomas, L. Rice, and D. White. 2019. The calculus concept inventory: a psychometric analysis and implications for use. *Int J Math Educ Sci Technol* 50:825–838.
10. Treagust, D. F. 1988. Development and use of diagnostic tests to evaluate students' misconceptions in science. *Int J Sci Educ* 10:159–169.
11. Huffman, D., and P. Heller. 1995. What does the Force Concept Inventory actually measure? *Phys Teach* 33:138–143. https://doi.org/10.1119/1.2344171.
12. Hake, R. 1998. Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66:64–74. https://doi.org/10.1119/1.18809.
13. Bloom, B. S. 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. 1st ed. Longmans, Green and Co., London.
14. Anderson, L. W., D. Krathwohl, P. Airasian, K. A. Cruikshank, R. E. Mayer, P. Pintrich, J. Raths, and M. C. Wittrock. 2001. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman, New York.
15. Soeharto, S., B. Csapó, E. Sarimanah, F. I. Dewi, and T. Sabri. 2019. A review of students' common misconceptions in science and their diagnostic assessment tools. *J Pendidikan* 8:247–266.
16. Driver, R. 1981. Pupils' alternative frameworks in science. *Eur J Sci Educ* 3:93–101.
17. Linke, R. D., and M. I. Venz. 1978. Misconceptions in physical science among non-science background students. *Res Sci Educ* 8:183–193.
18. Halloun, I., and D. Hestenes. 1985. Common sense concepts about motion. *Am J Phys* 53:1056–1065. https://doi.org/10.1119/1.14031.
19. Tamir, P. 1971. An alternative approach to the construction of multiple choice test items. *J of Biol Educ* 5:305–307.

https://thebiophysicist.kglmeridian.com | 2025-01-18

20. Hestenes, D., M. Wells, and G. Swackhamer. 1992. Force Concept Inventory. *Phys Teach* 30:141–158. https://doi.org/10.1119/1.2343497.
21. Thornton, R. K., and D. R. Sokoloff. 1998. Assessing student learning of Newton's laws: the force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *Am J Phys* 66:338–352. https://doi.org/10.1119/1.18863.
22. Korff, J. V., B. Archibeque, K. A. Gomez, T. Heckendorf, S. B. McKagan, E. C. Sayre, E. W. Schenk, C. Shepherd, and L. Sorell. 2016. Secondary analysis of teaching methods in introductory physics: a 50 k-student study. *Am J Phys* 84:969–974.
23. Fensham, P. J., J. Garrard, and L. West. 1981. The use of cognitive mapping in teaching and learning strategies. *Res Sci Educ* 11:121–129.
24. Halloun, I. A., and D. Hestenes. 1985. The initial knowledge state of college physics students. *Am J Phys* 53:1043–1055.
25. Marx, J. D. 2007. Normalized change. *Am J Phys* 75:87–91. https://doi.org/10.1119/1.2372468.
26. Miller, K., N. Lasry, O. Reshef, J. Dowd, I. Araujo, and E. Mazur. 2010. Losing it: the influence of losses on individuals' normalized gains. In Proceedings of the Physics Education Research Conference, Portland, OR, USA. 21 July 2010. American Institute of Physics. New York, NY, USA. pp. 229–232.
27. Beichner, R. 1994. Testing student interpretation of kinematics graphs. *Am J Phys* 62:750–762. https://doi.org/10.1119/1.17449.
28. Ding, L., R. Chabay, B. Sherwood, and R. Beichner. 2006. Evaluating an electricity and magnetism assessment tool: brief electricity and magnetism assessment. *Phys Rev Phys Educ Res* 2:010105. https://doi.org/10.1103/PhysRevSTPER.2.010105.
29. Landis, C., G. Lisenky, J. Lorenz, K. Meeker, C. Wamser, and A. Ellis. 2000. Chemistry Conceptests: A Pathway to Interactive Classrooms. Prentice Hall, Upper Saddle River, NJ.
30. Mulford, D. R., and W. R. Robinson. 2002. An inventory for alternate conceptions among first-semester general chemistry students. *J Chem Educ* 79:739–744.
31. Klymkowsky, M. W., and K. Garvin-Doxas. 2008. Recognizing student misconceptions through Ed's Tools and the Biology Concept Inventory. *PLOS Biol* 6:e3.
32. Williams, K. S., K. M. Fisher, D. L. Anderson, M. U. Smith, and J. E. Lineback. 2008. Using diagnostic test items to assess conceptual understanding of basic biology ideas: a plan for programmatic assessment. Conceptual Assessment in Biology Conference II. pp. 3–8. https://api.semanticscholar.org/CorpusID:58966893.
33. Odom, A. L., and L. H. Barrow. 1995. Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *J Res Sci Teach* 32:45–61.
34. Nehm, R. H., and L. Reilly. 2007. Biology majors' knowledge and misconceptions of natural selection. *BioScience* 57:263–272.
35. Tryon, G. S. 1980. The measurement and treatment of test anxiety. *Rev Educ Res* 50:343–372.
36. Hembree, R. 1988. Correlates, causes, effects, and treatment of test anxiety. *Rev Educ Res* 58:47–77.
37. United States Department of Health and Human Services. 2018. Protection of Human Subjects. Code of Federal Regulations, Government Printing Office.
38. Latham, D., J. Alper, and H. Sanabria. 2023. Biophysical Concept Inventory Survey. Accessed 3 February 2023. https://forms.gle/xsoHAQPyZe7XVFtm6.
39. American Council on Education. 2022. Carnegie Classification of Institutions of Higher Education. Accessed 23 August 2022. https://carnegieclassifications.acenet.edu/.
40. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J Mol Graph* 14:33–38.
41. Linn, M. C., E. Palmer, A. Baranger, E. Gerard, and E. Stone. 2015. Undergraduate research experiences: impacts and opportunities. *Science* 347:1261757.
42. Lopatto, D. 2010. Science in Solution: The Impact of Undergraduate Research on Student Learning. Research Corporation for Science Advancement, Tucson, AZ.
43. Coletta, V. P., and J. A. Phillips. 2005. Interpreting FCI scores: normalized gain, preinstruction scores, and scientific reasoning ability. *Am J Phys* 73:1172–1182.
44. Cochran, W. G. 1968. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 24:295–313. https://doi.org/10.2307/2528036.
45. Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Hillsdale, NJ.
46. Cohen, J. 1992. A power primer. *Psychol Bull* 112:155–159. https://doi.org/10.1037/0033-2909.112.1.155.
47. Kelley, K., and K. J. Preacher. 2012. On effect size. *Psychol Methods* 17:137–152. https://doi.org/10.1037/a0028086.
48. Sawilowsky, S. S. 2009. New effect size rules of thumb. *J Mod Appl Stat Methods* 8:597–599.
49. Wasserstein, R. L., and N. A. Lazar. 2016. The ASA statement on *p*-values: context, process, and purpose. *Am Stat* 70:129–133.
50. Nissen, J. M., R. M. Talbot, A. Nasim Thompson, and B. Van Dusen. 2018. Comparison of normalized gain and Cohen's *d* for analyzing gains on concept inventories. *Phys Rev Phys Educ Res* 14:010115.
51. Lestari, I. F. 2021. Experiential learning using STEM approach in improving students' problem solving ability. *J Phys Conf Ser* 1806:012005.
52. Khairati, K., W. Artika, M. A. Sarong, A. Abdullah, and H. Hasanuddin. 2021. Implementation of STEM-based experiential learning to improve critical thinking skills on ecosystem materials. *J Penelit Pendidikan* 7:752–757.
53. Pérez García, M., C. J. Luxford, T. L. Windus, and T. Holme. 2016. A quantum chemistry concept inventory for physical chemistry classes. *J Chem Educ* 93:605–612.
54. Andrich, D., I. Marais, and S. Humphry. 2012. Using a theorem by Andersen and the dichotomous Rasch model to assess the presence of random guessing in multiple choice items. *J Educ Behav Stat* 37:417–442.
55. Oberai, H., and I. M. Anand. 2018. Unconscious bias: thinking without thinking. *Hum Resour Manag Int Dig* 26:14–17. https://doi.org/10.1108/HRMID-05-2018-0102.
56. Dee, T., and S. Gershenson. 2017. Unconscious Bias in the Classroom: evidence and Opportunities. Stanford Center for Education Policy Analysis. https://cepa.stanford.edu/content/unconscious-bias-classroom-evidence-and-opportunities.
57. Lee, U. J., G. C. Sbeglia, M. Ha, S. J. Finch, and R. H. Nehm. 2015. Clicker score trajectories and concept inventory scores as predictors for early warning systems for large STEM classes. *J Sci Educ Technol* 24:848–860.
58. Middleton, F. 2019. The 4 Types of Validity in Research. Definitions & Examples. Accessed 16 March 2024. https://www.scribbr.com/methodology/types-of-validity/#:~:text=There%20are%20four%20main%20types%20of%20validity%3A%20Construct,test%20appear%20to%20be%20suitable%20to%20its%20aims%3F.

https://thebiophysicist.kglmeridian.com | 2025-01-18