



OPEN ACCESS

EDITED BY

Jodye I. Selco,
California State Polytechnic University,
Pomona, United States

REVIEWED BY

Stanley M. Lo,
University of California, San Diego,
United States
Michael J. Wolyniak,
Hampden–Sydney College, United States

*CORRESPONDENCE

Philip M. Reeves
✉ preeves1@jh.edu

RECEIVED 13 November 2023

ACCEPTED 11 March 2024

PUBLISHED 09 April 2024

CITATION

Reeves PM, Bauer M, Gill JC, Wang C,
Hanauer DI and Graham MJ (2024) More
frequent utilization of evidence-based
teaching practices leads to increasingly
positive student outcomes.
Front. Educ. 9:1337703.
doi: 10.3389/feduc.2024.1337703

COPYRIGHT

© 2024 Reeves, Bauer, Gill, Wang, Hanauer
and Graham. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

More frequent utilization of evidence-based teaching practices leads to increasingly positive student outcomes

Philip M. Reeves^{1*}, Melanie Bauer^{2,3}, Julia C. Gill³, Cong Wang⁴, David I. Hanauer⁵ and Mark J. Graham³

¹Master of Education in the Health Profession Program, Johns Hopkins University, Baltimore, MD, United States, ²Division of Research and Economic Development, Nova Southeastern University, Fort Lauderdale, FL, United States, ³Department of Ecology and Evolutionary Biology, STEM Program Evaluation and Research Lab, Yale University, New Haven, CT, United States, ⁴Department of Psychology, University of Michigan, Ann Arbor, MI, United States, ⁵Department of English, Indiana University of Pennsylvania, Indiana, PA, United States

Introduction: Evidence-based teaching (EBT) practices benefit students, yet our understanding of how frequently these strategies should be utilized in STEM courses is less established.

Methods: In this study, students ($n = 894$) of faculty who learned about how to implement EBTs from the Summer Institutes for Scientific Teaching were surveyed. The students rated the frequency of implementation of six types of EBTs after the course and completed a follow-up survey one year later to examine long-term outcomes.

Results: Class and student-level analyses indicated that students who reported being exposed to EBTs every couple of classes also reported that they learned more and had a higher ability to work in groups when compared to students who were exposed to EBTs less frequently.

Discussion: The results will help instructors and faculty development professionals understand how frequently EBTs should be incorporated when designing and modifying courses.

KEYWORDS

faculty development, STEM education, scientific teaching, course design, evidence based teaching practices

1 Introduction

There has been widespread effort over the last decade to improve the quality of teaching in undergraduate STEM courses ([American Association for the Advancement of Science, 2011](#); [President's Council of Advisors on Science and Technology, 2012](#)). Researchers have identified scientific teaching practices that relate to improved student outcomes. Examples of these practices include inclusive teaching, active learning, formative assessment, and the modeling of science reasoning and metacognitive strategies ([Handelsman et al., 2004](#)). Workshops and other faculty development opportunities have become increasingly available to teach faculty how to incorporate these practices into their instruction ([Simonson et al., 2022](#)). To help make the course modification process more manageable, faculty are often advised to make small, incremental changes to their courses ([Stein and Graham, 2020](#)). However, if these changes are too small, the new activities or teaching practices might not have a noticeable or immediate impact on student outcomes. Without positive feedback from students, faculty may prematurely abandon these teaching strategies and resist making subsequent modifications. The current study investigates the frequency with which

newly trained faculty were able to implement scientific teaching practices into their courses and provides initial guidance about the level of implementation needed to impact student outcomes.

2 Literature review

There is interest from teachers, researchers, administrators, and policymakers to improve learning outcomes and student retention in STEM fields (Handelsman et al., 2004; President's Council of Advisors on Science and Technology, 2012; Freeman et al., 2014; Bradforth et al., 2015). While there are several factors that can impact retention in STEM majors, offering high-quality teaching can have a dramatic, positive effect (Sithole et al., 2017). Research on evidence-based teaching practices such as active learning, formative assessment, and inclusive teaching has been shown to both increase engagement and the likelihood that students will remain in a science major (Hanauer et al., 2006; Auchincloss et al., 2014; Corwin et al., 2015). Nevertheless, implementing evidence-based teaching practices has been an enduring challenge in higher education—with up to 90% of STEM students citing concerns about teaching quality in their introductory courses (Seymour and Hewitt, 1997; Akiha et al., 2018). There is evidence to suggest that student dissatisfaction with teaching is at least partially due to incongruence between their expectations and their actual experience (Brown et al., 2017). For example, many first-year students are used to smaller high school classes that support active engagement with the content, peers, and instructors (Meaders et al., 2019). On the other hand, many introductory STEM courses in higher education still rely on passive lectures where students are more likely to report feeling isolated, dissatisfied, and disconnected from the content (Seymour and Hewitt, 1997; Cooper and Robinson, 2000).

Fortunately, STEM educators have recognized the importance of improving teaching quality in introductory STEM courses (American Association for the Advancement of Science, 2011; National Research Council, 2012; President's Council of Advisors on Science and Technology, 2012). An increasing number of colleges and universities host faculty development workshops to improve the quality of teaching. There are also national programs, such as the Summer Institutes for Scientific Teaching, that have provided instruction on how to implement evidence-based teaching practices. Initial evidence indicates that these faculty development efforts have: (a) positively impacted participants' motivation (Meixner et al., 2021); (b) increased pedagogical knowledge; (c) heightened the value faculty place on teaching (Palmer et al., 2016); (d) improved syllabi (Hershock et al., 2022); and subsequently improved their students' outcomes (Durham et al., 2020).

Despite the increasing prevalence of faculty development opportunities and the positive impact of these programs, implementation of evidence-based teaching practices still varies considerably across courses and instructors (Fairweather and Paulson, 2008; Durham et al., 2018). Participation in these programs, at most institutions, is voluntary. Faculty frequently cite misaligned incentives and time constraints as barriers to both attending training and fully implementing learner-focused instruction (Brownell and Tanner, 2012). Therefore, despite widespread efforts, lecture is still the most common instructional method used in introductory STEM courses (Stains et al., 2018).

In an effort to obtain faculty buy-in and to make the course (re) design process more manageable, faculty are often encouraged to make small changes over several semesters with the goal of incrementally improving their course over time (Stein and Graham, 2020). Nevertheless, gradual implementation may not produce immediate results. Without an

immediate impact on student outcomes or satisfaction, faculty may prematurely conclude that the new instructional methods are not effective in their specific learning context and revert to their prior pedagogical strategies. Therefore, it would be beneficial to understand the level of implementation that is necessary to see an impact on student outcomes. This would help faculty and faculty development professionals determine how to pace their initial modifications and increase the likelihood that both faculty and students see benefits.

Research that examines how frequently these teaching principles need to be applied within and across courses is still in its early stages. In a recent study, Reeves et al. (2023), found that utilizing evidenced-based teaching across multiple courses and instructors had a cumulative benefit to student learning and persistence in STEM. In that study, however, the frequency of implementation was not examined, as students were just asked to indicate if various evidence-based teaching practices occurred at all during the course. The purpose of this study is to examine how the frequency of implementation, as perceived by students, impacts student outcomes. Such findings would have the potential to reduce perceived (or real) barriers to initial implementation and encourage faculty to persist with the implementation of evidence-based teaching practices.

3 Materials and methods

3.1 Context and participants

The National Institute on Scientific Teaching offers a week-long professional development curriculum designed to develop college instructors' teaching ability in the following areas of “scientific teaching” active learning, formative assessment, and inclusivity and diversity. This curriculum is based on the Scientific Teaching Handbook by Handelsman et al. (2007). To date over 1,800 college faculty and instructors have attended a Summer Institutes workshop. A subset of these faculty and their students were recruited to participate in the current study. This project was granted exempt status from the Yale University IRB Human Subjects Committee.

3,142 students completed a survey the year after their instructor attended the summer institute. An additional follow-up survey was conducted a year later (2 years after the instructor attended) with 894 (28% of the original sample) responding. Two students in the follow-up survey responded relative to two different faculty members. The students rated 52 different instructors from 43 institutions. The number of students that responded per faculty member ranged from 2 to 80.

3.2 Procedures

3.2.1 Initial survey—implementation of teaching practices

Students reported their instructors' frequency of implementation for 26 evidence-based teaching practices (EBTs; see Appendix 1 for a list of these practices). The items were developed to assess teaching practices that were introduced at the Summer Institutes. Students indicated whether they experienced each teaching practice in the targeted course that semester. For the EBTs that they reported being exposed, participants reported how often they experienced each practice using the following scale: 1 (“once or twice”), 2 (“several times a semester”), 3 (“every couple of classes”), 4 (“about once a class”), or 5 (“multiple times per class”).

A Confirmatory Factor Analysis using LISREL 10.20 (Jöreskog and Sörbom, 2019) was conducted to evaluate the extent to which the predetermined structure (Graham, 2019) was a good fit for the observed data. A full description of the factor analyses and the model fit indices are provided in Appendix 2. The analysis provides evidence to support the use of the six general categories of EBT.

Active learning (6 items, $\alpha=0.71$) items considered activities that promoted student engagement in learning activities. *Assessment* (3 items, $\alpha=0.52$) items included activities that provided formative feedback to students during class. *Diversity* (3 items, $\alpha=0.75$) items focused on how the instructor communicated that all students could succeed in the class. *Inclusive teaching* (4 items, $\alpha=0.77$) items addressed the representation of multiple perspectives within the course. *Metacognition* (4 items, $\alpha=0.75$) items measured the use and modeling of learning strategies. *Science reasoning* (6 items, $\alpha=0.78$) items considered opportunities to participate in the scientific process. These EBTs are defined in Table 1.

3.2.2 Follow-up survey—student outcomes

One year later, students were asked to reflect on the impact of the targeted course. Students were asked to indicate if 15 outcome items were true (scored as 0=False and 1=True). The items were drafted after commonly reported desired outcomes were identified in the Scientific Teaching handbook (on which the instructor participants were trained; Handelsman et al., 2007) and the National Research Council (2012). The items generally examined students' interest in learning, awareness of the purpose of the course, interest in working in groups, interest in continuing in STEM, and understanding research processes.

To categorize the outcomes measured on the 15 checklist items, we conducted a series of exploratory factor analyses. An exploratory factor analysis was used because the factor structure of these items had not previously been examined. Item Response Theory (IRT) was utilized because these items are dichotomous rather than continuous. After checking the factor loadings of each item, we found that item 4 (i.e., More aware of how I learn best) had low factor loadings (<0.30)

on all potential factors, and item 8 (i.e., More interested in science) was cross-loaded on two factors. We decided to eliminate these two items from the following analyses and re-ran the factor analysis. Model fit indices of all tested models are presented in Appendix 3. The resulting structure included four factors.

Course objectives (5 items, $\alpha=0.73$) that measured perceived achievement of general course objectives. *Group work* (2 items, $\alpha=0.72$) that focused on the perceived ability to work in groups. *Science career* (3 items, $\alpha=0.74$) that concentrated on the likelihood of persisting in a STEM field. *Research* (3 items, $\alpha=0.68$) that centered on perceived ability to conduct research.

4 Analyses and results

Two types of analyses were conducted. The first type of analysis looked at EBT exposure at the individual level and the second was conducted at the classroom level.

4.1 Individual-level

4.1.1 Clustering students based on the implementation of EBTs

Due to variance in the students' perception of EBT utilization within classrooms, the data was analyzed at the individual level. In other words, we grouped students based on the frequency of their perceived exposure regardless of the instructor. A two-step cluster analysis in SPSS version 25. Subscale scores were entered in the analysis, which identified two groups with a fair silhouette. 429 students perceived high levels of implementation and 465 students perceived a low level of implementation. Figure 1 shows the mean of each cluster with more detailed descriptives in Appendix 4.

TABLE 1 Evidence based teach practices.

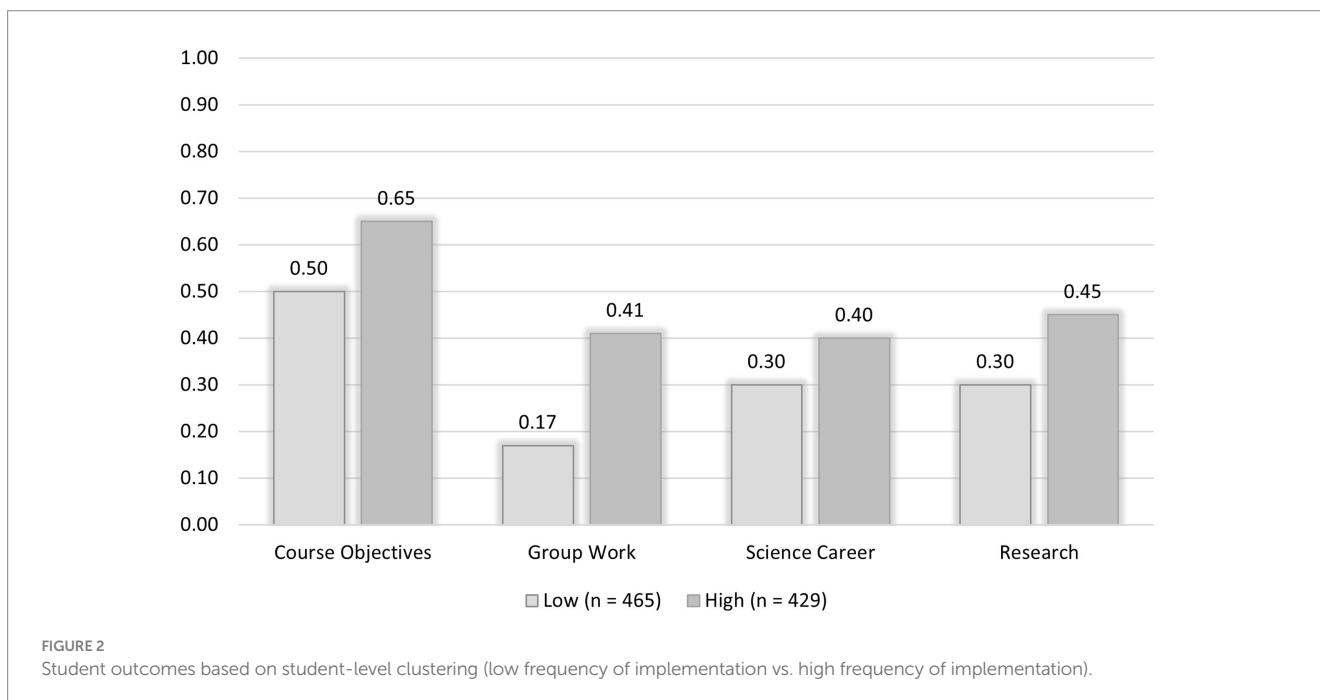
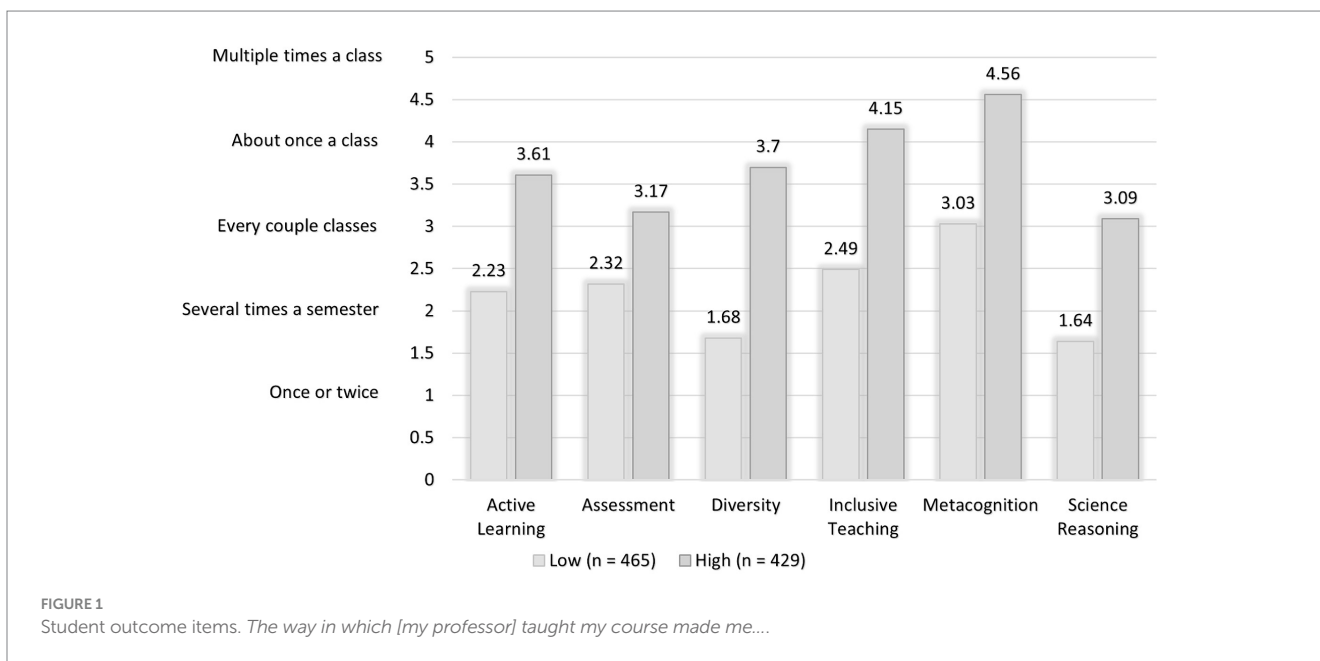
Term	Definition	Examples
Active learning	Providing students with the opportunity to discover and apply concepts on their own rather than passively receiving the information in a lecture (Prince, 2004).	Discussion, peer collaboration, experimentation, and problem solving.
Formative Assessment	Providing feedback while learning is occurring (Bennett, 2011).	Informal oral questioning, think-pair-share, peer assessment, clicker polls, quizzes, one-minute essays, homework, and prompts that promote student reflection.
Diversity and Collaboration	Providing opportunities to participate and learn in diverse groups (Singer et al., 2020).	Authentic learning experiences, jigsaw, simulations, case studies, guided discussion, and problem-based learning activities.
Inclusive Teaching	Creating equitable, accessible, and welcoming environments by responding to individual differences (Dewsbury and Brame, 2019).	Incorporating diverse authors, viewpoints, and examples during lessons; considering avenues for making course materials more affordable and accessible to all students; developing individual education plans; and encouraging and providing opportunities for equitable participation in class activities.
Metacognitive Learning Strategies	Providing opportunities for students to plan, monitor, and assess their own thinking and learning (Medina et al., 2017).	Journaling, reflective questioning, cognitive modeling, think alouds, peer mentoring
Science Reasoning	Providing students the opportunity to understand the world through hypothesis, experimentation, and evaluation (Osborne, 2013).	Designing experiments, troubleshooting experiments, critiquing research studies, generating hypotheses

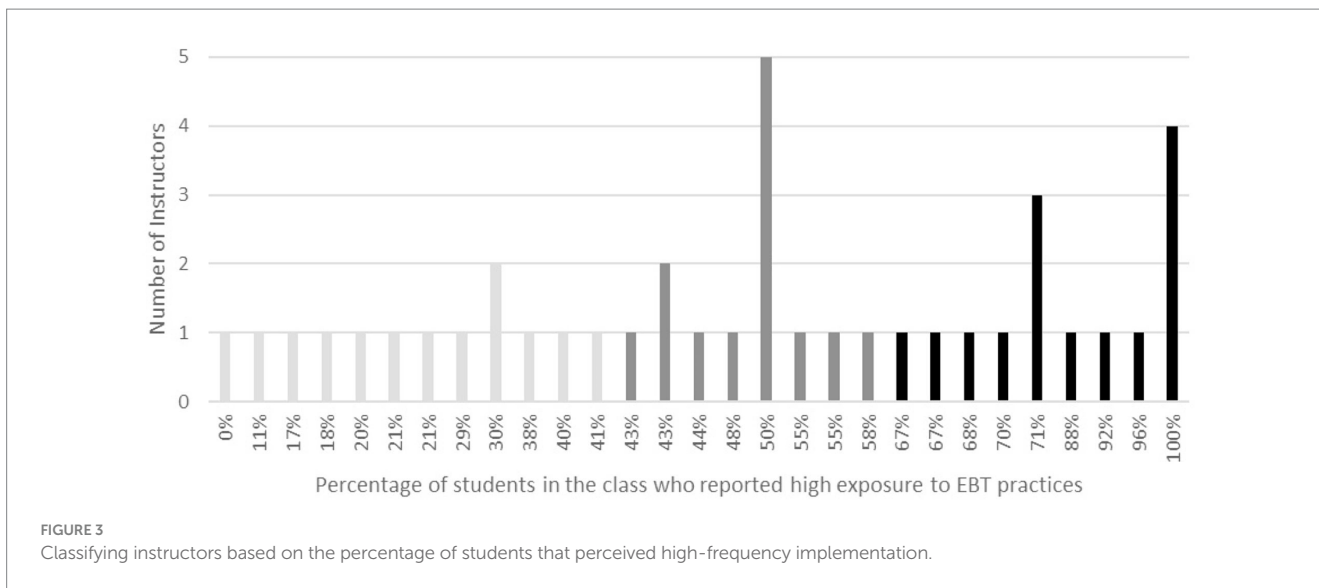
4.1.2 Confirming individual perceptions of implementation frequency

To verify the differences between perceptions of EBT implementation a series of independent t-tests were calculated to determine if the groups of students differed on the six EBT factors. A statistically significant difference was found on all factors [active learning ($t(801.94) = 24.41, p < 0.001$); assessment ($t(797.97) = 14.75, p < 0.001$); diversity ($t(676.82) = 28.15, p < 0.001$); inclusive teaching ($t(808.26) = 22.95, p < 0.001$); metacognition ($t(889.47) = 22.72, p < 0.001$); and scientific reasoning ($t(756.30) = 29.39, p < 0.001$)]. This suggests that a high level of EBT occurs at least every couple of classes (see Figure 1).

4.1.3 Follow up on student outcomes based on classroom-level student-level clustering

A logistic regression equation was used to predict implementation level (high vs. low) using the subscale scores for the four outcomes as independent variables. Means are presented in Figure 2 and more detailed descriptive statistics are found in Appendix 5. The logistic regression equation was statistically significant, $X^2(4) = 102.39, p < 0.001$. The model explained 14.4% of variance (Nagelkerke R²). Group work [$X^2(1) = 44.91, p < 0.001, OR = 3.81, CI = 2.58-5.63$] and course objectives [$X^2(1) = 12.61, p < 0.001, OR = 2.57, CI = 1.53-4.33$] were found to be statistically significant predictors.





4.2 Classroom level

4.2.1 Determining the level of EBT implementation

From the analysis of the individual-level descriptive statistics, it became apparent that student perceptions of EBT implementation varied (sometimes) considerably within the same classroom, despite theoretically receiving the same instruction. To account for the nested nature of the data, the second method was conducted at the whole classroom level. In other words, the entire classroom was classified as having received a high, medium, or low level of EBT exposure.

To classify instructors as being low or high implementers of EBTs, the percentage of students per classroom that perceived their instructor to be a high-frequency implementer (i.e., students categorized in the high category in the individual level cluster analysis) was calculated. After examining the distribution (Figure 3) there was no clear-cut point between high and low-implementing instructors. Therefore, in order to categorize the instructors, the instructors were divided into three equal categories based on the percentage of students that reported high exposure to EBTs (see Table 2). The top third of instructors ($n = 14$) were classified as being high-implementing classrooms. These instructors had more than 58% of students reporting high exposure to EBTs. The bottom third of instructors ($n = 13$) had less than 43% of students reporting high exposure to EBTs. The middle third ($n = 13$) was eliminated from subsequent analyses due to potential measurement error around the cut scores.

4.2.2 Confirming classroom implementation category

To confirm the differences in EBT implementation between the classroom categories, we conducted independent t -tests. Figure 4 shows the mean scores of the six EBT factors for each classroom category. The average from all students in the high implementers' classrooms reported significantly higher scores on all six EBT factors than all students in the low implementers' classrooms [active learning ($t(25) = 4.50, p < 0.001$); assessment ($t(25) = 2.49, p = 0.02$); diversity

TABLE 2 Number of classrooms by EBT implementation level.

Instructor implementation category	Percentage of students in the class who reported high exposure to EBT practices	Number of Instructors
Low	Less than 43%	13
Medium	Between 43–58%	13
High	More than 58%	14

($t(25) = 2.83, p = 0.001$); inclusive teaching ($t(25) = 2.82, p = 0.001$); metacognition ($t(25) = 3.10, p = 0.005$); and scientific reasoning ($t(25) = 5.67, p < 0.001$). The data suggests that in high-implementing classrooms EBTs occurred several times a semester or every couple of classes. Students in the low-implementing classrooms only reported EBTs occurring once or twice a semester. Descriptive statistics can be found in Appendix 6.

4.2.3 Follow up on student outcomes based on the classroom-level classification

To understand how students are impacted by the implementation level of EBT, we conducted independent t -tests with each of the four outcome factors. Figure 5 shows the mean scores of these outcomes for each level of classroom implementation level. Students in the high implementers' classrooms reported significantly higher scores on all outcome factors than students in the low implementers' classrooms [course objectives ($t(25) = 2.48, p = 0.02$); group work ($t(25) = 4.39, p < 0.001$); science career ($t(25) = 2.29, p = 0.03$); research ($t(20.17) = 2.65, p = 0.02$)]. Descriptive statistics can be found in Appendix 7.

5 Discussion

The study examined how the level of EBT implementation impacts student outcomes. A majority of instructors who completed the training at the summer institute utilized each EBT practice (e.g.,

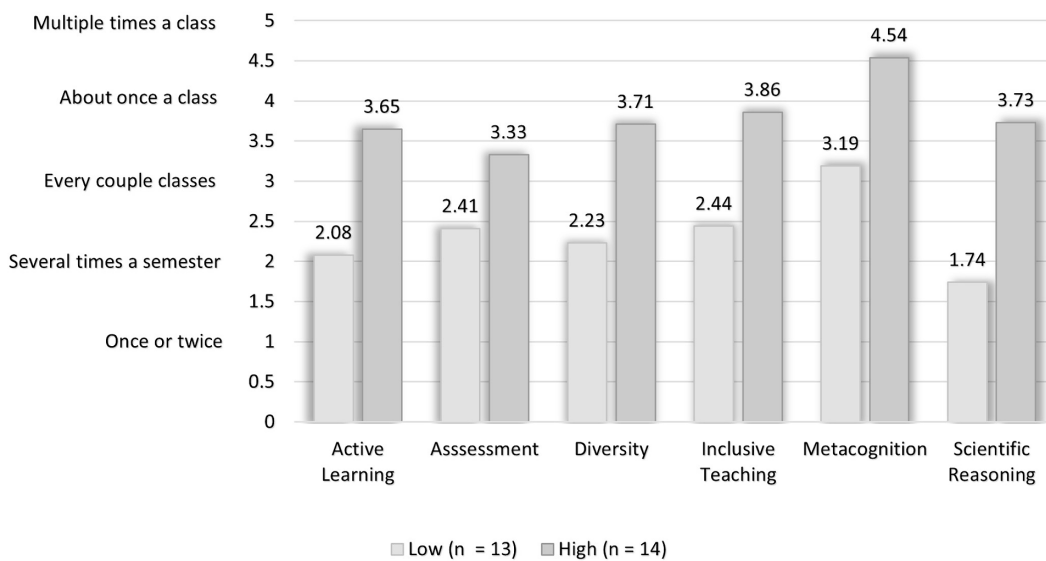


FIGURE 4
A comparison of student outcomes based on classroom classification (low frequency of implementation vs. high frequency of implementation) of EBT implementation.

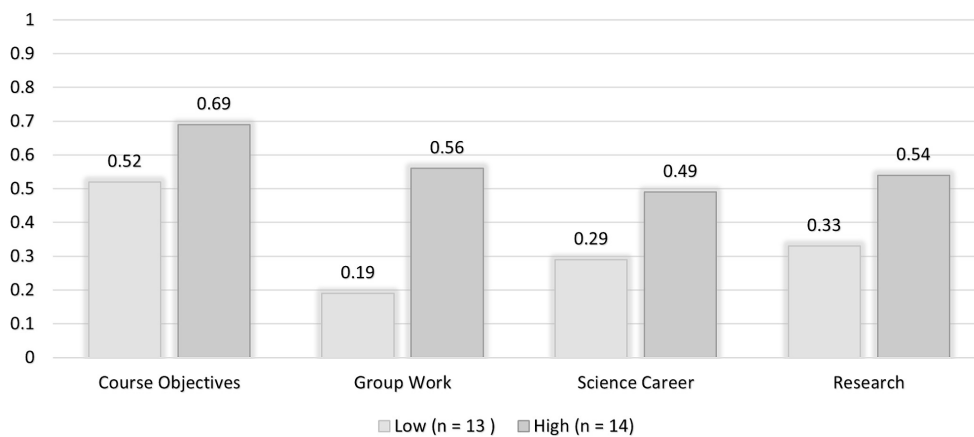


FIGURE 5
A comparison of the average frequency of EBT implementation based on the classroom-level classification (low frequency of implementation vs. high frequency of implementation).

active learning) several times a semester. This is consistent with advice given by faculty developers to make gradual incremental changes to instructional methods (Stein and Graham, 2020). Nevertheless, small degrees of implementation may not produce noticeable differences in student outcomes. The results from both analyses—classroom level and student level – in this study suggest that the perceived frequency of EBT implementation is associated with student outcomes. Including EBTs every couple of classes relates to more positive student outcomes when compared to only including these practices several times across the semester. Intuitively, the result makes sense as it follows the adage “more is better,” but also implies that instructors may need to make more sudden or substantial changes to courses that primarily relied on lectures and teacher-directed learning activities.

The population of instructors in this study were predominantly employed at large research institutions, which reflects the participants at the Summer Institutes. Faculty at these institutions may have different pressures (i.e., rewarded for focusing more on research than teaching) and be more likely to teach larger classes, particularly at the introductory level when compared to faculty from teaching-focused colleges (Brownell and Tanner, 2012). These factors could influence the time that faculty can dedicate to instructional design and building trust with their students (Wang et al., 2021). However, the results suggest that it is important for these faculty to dedicate time to including EBTs regularly throughout the course. Including more EBTs may help mitigate student dissatisfaction with learning expectations that can be common in large classes by making learners feel more engaged in the content (Seymour and Hewitt, 1997; Cooper and

Robinson, 2000). Future studies should examine whether the impact of the frequency of EBT implementation varies based on class size or institutional teaching practices. It is possible that small class sizes and the overall level of an institution's adoption of EBTs could impact the amount of implementation needed to impact student performance within a particular course. However, it is also possible that there is a level of implementation that has diminishing returns (e.g., students get tired of reflecting everyday).

The study had several limitations. First, the implementation of EBTs by instructors was reported by students. The students may not have been fully aware of when instructors were trying to utilize some techniques and could have inaccurately perceived, or remembered, the frequency with which they were incorporated into the course. Measuring frequency with a Likert scale may have mitigated the impact of these inaccurate perceptions, though it also limits the precision of the measurement and any subsequent instructional recommendations. Future studies would benefit from taking an observational approach to measuring implementation. If done in conjunction with student surveys, this would also help researchers investigate the variability in student reporting of EBTs within classrooms.

Second, the outcome data was also self-reported by students. The factor analyses provided evidence that the subscales are measuring unique concepts. However, this result could be strengthened with the inclusion of knowledge tests or other indicators of student achievement in future studies. Finally, the faculty who were recruited for this study attended the Summer Institutes for Scientific Teaching, which may signify a higher level of interest in pedagogy and instructional design, when compared to other faculty. Successful implementation of EBTs likely varies based on the motivation of faculty (Richardson et al., 2014). Future studies would benefit from examining how a gradual course design process interacts with faculty motivation for teaching. Some of the faculty in this study also may have included EBTs in their instruction before their training, so we cannot be sure of how many modifications were made to courses because of the summer workshop. This information would be useful for making more specific recommendations on how many changes to instruction should be made at any given time.

Despite the limitations, these findings present initial evidence to suggest that more frequent utilization will lead to larger benefits, and incorporating EBTs every couple of classes will likely lead to a more noticeable impact. Prior to this study, the impact of the frequency of EBT had not been examined. Future studies should evaluate faculty perceptions of their implementation to determine if the frequency of EBT use impacts their perception of student performance, teaching self-efficacy, job satisfaction, and intention to continue modifying their instruction to include more EBTs. A range of optimal implementation likely exists for each teaching practice. This study provides evidence that infrequent implementation does not lead to maximum benefits. Additional insight into these variables would allow researchers and faculty developers to provide better guidance on how to pace and scale course modification in a manner that benefits both faculty and students.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Yale University IRB Human Subjects Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because it was designated as exempt survey research.

Author contributions

PR: Writing – original draft, Writing – review & editing. MB: Writing – original draft. JG: Writing – review & editing. CW: Writing – original draft. DH: Writing – review & editing. MG: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported through a National Science Foundation Transforming Undergraduate Research in the Sciences (TUES) grant (NSF no. 1323258).

Acknowledgments

We thank the course instructors and students who participated in the project. We thank our collaborators on this grant, including Jane Cameron Buckley, Andrew Cavanaugh, Brian Couch, Mary Durham, Jennifer Frederick, Monica Hargraves, Claire Hebbard, Jennifer Knight, and William Trochim. Support for the Summer Institutes on Scientific Teaching was made possible through funding from the Howard Hughes Medical Institute originally awarded to Jo Handelsman.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2024.1337703/full#supplementary-material>

References

- Akihi, K., Brigham, E., Couch, B. A., Lewin, J., Stains, M., Stetzer, M. R., et al. (2018). Frontiers Media SA). What types of instructional shifts do students experience? Investigating active learning in science, technology, engineering, and math classes across key transition points from middle school to the university level. *Front. Educ.* 2:68. doi: 10.3389/feduc.2017.00068
- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: a call to action*. Washington, DC: American Association for the Advancement of Science.
- Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., et al. (2014). Assessment of course-based undergraduate research experiences: a meeting report. *CBE Life Sci. Educ.* 13, 29–40. doi: 10.1187/cbe.14-01-0004
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assess. Educ. Princ. Policy Pract.* 18, 5–25. doi: 10.1080/0969594X.2010.513678
- Bradforth, S. E., Miller, E. R., Dichtel, W. R., Leibovich, A. K., Feig, A. L., Martin, J. D., et al. (2015). University learning: improve undergraduate science education. *Nature* 523, 282–284. doi: 10.1038/523282a
- Brown, T. L., Brazeal, K. R., and Couch, B. A. (2017). First-year and non-first-year student expectations regarding in-class and out-of-class learning activities in introductory biology. *J. Microbiol. Biol. Educ.* 18, 1–9. doi: 10.1128/jmbe.v18i1.1241
- Brownell, S. E., and Tanner, K. D. (2012). Barriers to faculty pedagogical change: lack of training, time, incentives, and... tensions with professional identity? *CBE Life Sci. Educ.* 11, 339–346. doi: 10.1187/cbe.12-09-0163
- Cooper, J. L., and Robinson, P. (2000). The argument for making large classes seem small. *New Dir. Teach. Learn.* 2000, 5–16. doi: 10.1002/tl.8101
- Corwin, L. A., Graham, M. J., and Dolan, E. L. (2015). Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. *CBE Life Sci. Educ.* 14:es1. doi: 10.1187/cbe.14-10-0167
- Dewsbury, B., and Brame, C. J. (2019). Inclusive teaching. *CBE Life Sci. Educ.* 18:fe2. doi: 10.1187/cbe.19-01-0021
- Durham, M. F., Aragón, O. R., Bathgate, M. E., Bobrownicki, A., Cavanagh, A. J., Chen, X., et al. (2020). Benefits of a college STEM faculty development initiative: instructors report increased and sustained implementation of research-based instructional strategies. *J. Microbiol. Biol. Educ.* 21, 21–22. doi: 10.1128/jmbe.v21i2.2127
- Durham, M. F., Knight, J. K., Bremers, E. K., DeFreeze, J. D., Paine, A. R., and Couch, B. A. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the measurement instrument for scientific teaching-observable (MISTO). *Int. J. STEM Educ.* 5, 31–15. doi: 10.1186/s40594-018-0128-1
- Fairweather, J. S., and Paulson, K. (2008). “The evolution of American scientific fields: disciplinary differences versus institutional isomorphism (197–212)” in *Cultural perspectives on higher education*. eds. J. Välimaa and O. H. Ylijoki (Dordrecht: Springer)
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc. Nat. Acad. Sci. USA* 111, 8410–8415. doi: 10.1073/pnas.1319030111
- Graham, M. J. (2019). *Impact of the Summer Institutes on faculty teaching and student learning. Gordon conference on undergraduate biology education research*, Bates College, ME.
- Hanauer, D. I., Jacobs-Sera, D., Pedulla, M. L., Cresawn, S. G., Hendrix, R. W., and Hatfull, G. F. (2006). Teaching scientific inquiry. *Science* 314, 1880–1881. doi: 10.1126/science.1136796
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., et al. (2004). Scientific teaching. *Science* 304, 521–522. doi: 10.1126/science.1096022
- Handelsman, J., Miller, S., and Pfund, C. (2007). *Scientific Teaching (2007)* (W. H. Freeman, New York).
- Hershock, C., Pottmeyer, L. O., Harrell, J., le Blanc, S., Rodriguez, M., Stimson, J., et al. (2022). Data-driven iterative refinements to educational development services: directly measuring the impacts of consultations on course and syllabus design. *Imp. Acad. J. Educ. Dev.* 41:9. doi: 10.3998/tia.926
- Jöreskog, K. G., and Sörbom, D. (2019). LISREL 10.20 [Computer Software]. Scientific Software International.
- Meaders, C. L., Toth, E. S., Lane, A. K., Shuman, J. K., Couch, B. A., Stains, M., et al. (2019). “What will I experience in my college STEM courses?” an investigation of student predictions about instructional practices in introductory courses. *CBE Life Sci. Educ.* 18:ar60. doi: 10.1187/cbe.19-05-0084
- Medina, M. S., Castleberry, A. N., and Persky, A. M. (2017). Strategies for improving learner metacognition in health professional education. *Am. J. Pharm. Educ.* 81:78. doi: 10.5688/ajpe81478
- Meixner, C., Altman, M., Good, M., and Ward, E. (2021). Longitudinal impact of faculty participation in a course design institute (CDI): faculty motivation and perception of expectancy, value, and cost. *Imp. Acad. J. Educ. Dev.* 40. doi: 10.3998/tia.959
- National Research Council (2012). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. Washington D.C.: The National Academies Press.
- Osborne, J. (2013). The 21st century challenge for science education: assessing scientific reasoning. *Think. Skills Creat.* 10, 265–279. doi: 10.1016/j.tsc.2013.07.006
- Palmer, M. S., Streifer, A. C., and Williams-Duncan, S. (2016). Systematic assessment of a high-impact course design institute. *Imp. Acad.* 35, 339–361. doi: 10.1002/tia2.20041
- President’s Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: U.S. Government Office of Science and Technology (pp. 36–38).
- Prince, M. (2004). Does active learning work? A review of the research. *J. Eng. Educ.* 93, 223–231. doi: 10.1002/j.2168-9830.2004.tb00809.x
- Reeves, P. M., Cavanagh, A. J., Bauer, M., Wang, C., and Graham, M. J. (2023). Cumulative cross course exposure to evidence-based teaching is related to increases in STEM student buy-in and intent to persist. *Coll. Teach.* 71, 66–74. doi: 10.1080/87567555.2021.1991261
- Richardson, W., Karabenick, S., and Watt, H. M. G. (2014). *Teacher motivation: theory and practice*. New York: Routledge.
- Seymour, E., and Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview.
- Simonson, S. R., Earl, B., and Frary, M. (2022). Establishing a framework for assessing teaching effectiveness. *Coll. Teach.* 70, 164–180. doi: 10.1080/87567555.2021.1909528
- Singer, A., Montgomery, G., and Schmoll, S. (2020). How to foster the formation of STEM identity: studying diversity in an authentic learning environment. *Int. J. STEM Educ.* 7, 1–12. doi: 10.1186/s40594-020-00254-z
- Sithole, A., Chiyaka, E. T., McCarthy, P., Mupinga, D. M., Bucklein, B. K., and Kibirige, J. (2017). Student attraction, persistence and retention in STEM programs: successes and continuing challenges. *High. Educ. Stud.* 7, 46–59. doi: 10.5539/hes.v7n1p46
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., et al. (2018). Anatomy of STEM teaching in north American universities. *Science* 359, 1468–1470. doi: 10.1126/science.aap8892
- Stein, J., and Graham, C. R. (2020). *Essentials for blended learning: a standards-based guide*. New York, NY: Routledge.
- Wang, C., Cavanagh, A. J., Bauer, M., Reeves, P. M., Gill, J. C., Chen, X., et al. (2021). A framework of college student buy-in to evidence-based teaching practices in STEM: the roles of trust and growth mindset. *CBE Life Sci. Educ.* 20:ar54. doi: 10.1187/cbe.20-08-0185