What really impacts the use of active learning in undergraduate STEM education? Results from a national survey of chemistry, mathematics, and physics instructors.

Naneh Apkarian[1¶], Charles Henderson[2¶*], Marilyne Stains[3&], Jeffrey R. Raker[4&], Estrella Johnson[5&], Melissa H. Dancy[6&]

[1] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, Arizona, USA
[2] Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, Michigan, USA
[3] Department of Chemistry, University of Virginia, Charlottesville, Virginia, USA
[4] Department of Chemistry, University of South Florida, Tampa, Florida, USA
[5] Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA
[6] Center for STEM Learning, University of Colorado Boulder, Boulder, Colorado, USA

* Corresponding author

E-mail: charles.henderson@wmich.edu (CH)

[¶] NA and CH are Joint Senior Authors and contributed equally to this work.

[&] These authors contributed equally to this work.

Abstract

Six common beliefs about the usage of active learning in introductory STEM courses are investigated using survey data from 3769 instructors. Three beliefs focus on contextual factors: class size, classroom setup, and teaching evaluations; three focus on individual factors: security of employment, research activity, and prior exposure. The analysis indicates that instructors in all situations can and do employ active learning in their courses. However, with the exception of security of employment, trends in the data are consistent with beliefs about the impact of these factors on usage of active learning. Implications of these results for institutional and departmental policies to facilitate the use of active learning are discussed.

Introduction

Undergraduate science, technology, engineering, and mathematics (STEM) instruction in the United States has faced increasing scrutiny for years, with particular focus on introductory courses. General Chemistry, Introductory Quantitative Physics, and Single-Variable Calculus are three examples of gatekeeper introductory courses - they are high enrollment, high-risk, foundational courses that have outsized impact on students' pathways to a STEM major [1]. Low passing rates in these courses have drawn much attention, but there is evidence to suggest that negative learning experiences dominate students' reasons for leaving. The percentage of STEM-intending students who complete an undergraduate STEM degree has stayed at roughly 40% since 1997, despite an increasing demand for scientists and technicians with a bachelor's degree [2,3]. While increasing numbers of women and students of color enter STEM majors they continue to leave at high rates, indicating a continued and substantial loss of natural talent and interest in the sciences [3]. These persistent concerns have drawn attention from university administrators, researchers, and governing bodies seeking to increase the pool of STEM graduates for economic and social reasons [4].

Decades of research in mathematics and science education has led to the development of active learning instructional strategies that are empirically demonstrated to promote content understanding, attitudes, and retention among all students, and to reduce achievement gaps between dominant and underrepresented groups in STEM [3,5,6]. In the broadest sense, *active learning* refers to classroom strategies which move away from a transmission or "telling" model (the classic "lecture") toward a model where students actively engage in problem-solving and knowledge creation [5,6]. Specific strategies might leverage individual investigations, team-based problem-solving, and/or whole class discussions. Even when broadly conceived, the use of such strategies in undergraduate STEM courses remains sparse despite concerns about student success rates, and despite increasing awareness of the need to implement active learning [3,7,8]. This paper examines empirical data related to six common beliefs about factors that impact instructors' use of active learning [9,10]. Three of these beliefs are primarily about contextual factors:

1. Large class sizes hinder the use of active learning.
2. Traditional fixed-seat classrooms hinder the use of active learning.
3. Emphasizing student evaluations of teaching hinders the use of active learning.

Three are primarily individual instructor characteristics:

4. Not having security of employment (e.g., tenure) hinders the use of active learning.
5. High levels of research activity hinders the use of active learning.
6. Experience with active learning as a student, or as a student instructor, facilitates the use of active learning.

We selected these beliefs in part because each can be linked to policy decisions and institutional priority setting. For example with the contextual factors: directing institutional funds toward decreasing class sizes by hiring additional faculty; building new classrooms designed to facilitate group work as opposed to new auditorium-style lecture halls; or revising assessments of teaching and their use in professional review. With regards to the individual characteristics, holding these beliefs can impact who is targeted by a change initiative, perhaps focusing on newer faculty and instructors or creating factions of teaching-focused and research-focused faculty. Despite their potential impact on important decisions and priority-setting, strong empirical data related to these beliefs has not previously been available.

Methods

Data for this report comes from a national survey of postsecondary instructors teaching introductory STEM courses at two-year colleges (TYC), four-year colleges (predominantly undergraduate institutions, PUI), and universities (UNI) in the United States. Data collection was conducted in Spring 2019, and the final sample reported on here consists of 3,769 respondents who were primary instructors of a general chemistry, single-variable calculus, or introductory quantitative physics course in the 2017-18 or 2018-19 academic year.

Data collection

A new survey instrument was developed by the six authors for this project. The full survey covered five main topics: (1) course context and details; (2) instructional practice; (3) awareness and usage of active learning instruction; (4) perceptions, beliefs, and attitudes related to students, learning, and departmental context; (5) personal demographics and experience. Specific items and the overall format were informed by previous large studies of chemistry [8,11,12], mathematics [13,14], and physics education [15,16]. The exact wording of the survey items included in this analysis are included in the supplement. When possible, previously validated instruments and scales were reproduced in their entirety as part of the survey. A web-based version of the instrument was built and distributed in partnership with the American Institute of Physics Statistical Research Center. Over 18,000 people were invited to participate, a roster compiled from publicly available information and communication with department chairs by members of the American Institute of Physics Statistical Research Center. Invitations were sent via email, with follow-up reminders sent to those who had not opened the survey at roughly two-week intervals over the course of six weeks.

The initial page of the survey served to inform participants of the nature of the study, their involvement, and potential risk. Informed consent was collected digitally, in accordance with the Institutional Review Board of Western Michigan University policies. The informed consent was followed by eligibility screening questions to ensure that those who participated had been the primary instructor of a general chemistry, single-variable calculus, or introductory quantitative physics course in the 2017-18 and/or 2018-19 academic years that was not taught entirely online. At the end of the data collection period, responses were reviewed and any participants who had not filled out a single post-eligibility question were removed as non-respondents. This resulted in a data set including 3,769 individuals. Table 1 presents some information about the range of participants included in this study, indicating that the responses are not heavily weighted toward a particular discipline or institution type and allowing for comparison to particular contexts.

Table 1. Table of respondents by institution type, discipline, and rank. Institution type and discipline were ascertained by the research team when the survey roster was developed. Rank was reported by participants and not required, hence the discrepancies in the total; proportion is out of those who provided a response.

Group	*Count*	*Proportion*	*Group*	*Count*	*Proportion*
Institution Type			**Rank**		
University (UNI)	1541	0.41	Professor	1052	0.33
Predominantly undergraduate institution (PUI)	1129	0.30	Assoc. Professor	692	0.21
Two-year college (TYC)	1099	0.29	Asst. Professor	543	0.17
Total	*3769*		Lecturer	773	0.24
Discipline			Visiting	102	0.03
Chemistry	1244	0.33	Postdoc	25	0.01
Mathematics	1349	0.36	Grad Student	44	0.01
Physics	1176	0.31	*Total*	*3231*	
Total	*3769*		No response	538	-

Data analysis

We conducted analyses to understand to what extent this survey data were consistent with each belief. The beliefs we investigate center on the usage of active learning, which we measure via instructor self-report of how class time is spent [11,17]. In particular, we use the reported percentage of class time spent in lecture as a proxy for the percentage of class time spent in *non-active learning* activities, following others in lumping together various active learning strategies for comparison to traditional straight lecture [5,6]. After removing responses which did not report on 100% of class time, 3641 survey responses were retained for analysis. Other data used in these analyses (e.g., course enrollment or research activity) also come from self-report survey data, and additional data reduction was performed on a case-by-case basis to omit non-response data, hence the changing N values reported in the findings. Further details of the methods, including data cleaning, data reduction, and tables of statistical results are included in the supplementary materials. For each belief, we present relevant data, analyses, and practical interpretations. In the final section, we summarize the results and suggest implications for policy makers and change agents hoping to increase the use of active learning instruction in introductory STEM courses.

Findings: Contextual factors

Course enrollment

One commonly espoused barrier to the use of active learning instructional strategies, particularly those involving student-student engagement, is large class sizes [9,10]. Participants reported the typical enrollment of their course, and these responses were binned into six size categories: 0-19; 20-29; 30-39; 40-59; 60-99; and 100 or more. Responses in the form of a range (e.g., "25-40") were averaged and binned according to that average (see supplemental information for additional detail). ANOVA indicates a small-to-medium effect [18,19] of class size on percentage of class time spent in lecture (Fig 1). From post hoc testing, the largest classes (those with 100+ students) have the highest percentage of lecture.

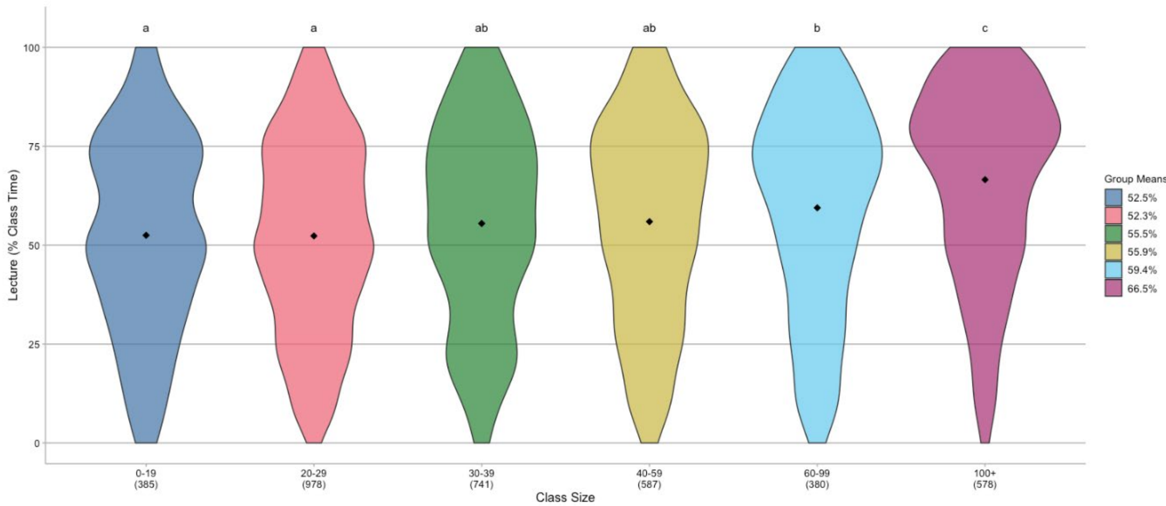

Fig 1. Violin plots of lecture (as percentage of class time) reported by instructors with different class sizes. Group means are indicated in the plots and reported in the legend; N is indicated below each category on the horizontal axis; common letters indicate group means that are not statistically different in Tukey HSD post hoc testing at a 95% confidence level. Class size is a main effect on lecture, $F(5,3630) = 27.9$, $p < 0.001$, $\eta^2 = 0.04$ (small/medium).

The variation we observe in this data suggests that there are instructors in all class size groupings that use active learning, and that there are instructors in all class size groupings that primarily lecture. However, the analysis suggests there is more lecture in the largest classes. For change agents, this suggests that decreasing class size, particularly avoiding offering very large courses, would support the use of active learning in introductory STEM courses. Another implication of these findings is that, while lecturing continues to be a norm in introductory STEM, there are a substantial number of instructors who use little lecture even with large classes[1]. Thus, it is possible to use active learning in large classes. There exist research-based instructional strategies that are particularly well-suited to increasing student-student engagement in large courses, such as Peer Instruction [20,21]. Helping instructors select and implement appropriate instructional strategies for their class size is an important way to increase the use of active learning.

[1] Details of non-lecture activities are included in the supplemental material.

Classroom setup

Traditional classrooms are also referenced as a barrier to implementing active learning strategies [10]. Participants in our study indicated whether their classroom was designed to accommodate group work (e.g., tables, movable desks) or were more traditional lecture halls with fixed seats/desks. There is a medium effect [18,19] on the amount of lecture reported by instructors in the two groups, with more lecture reported by those teaching in traditional fixed-seat classrooms (Fig 2).

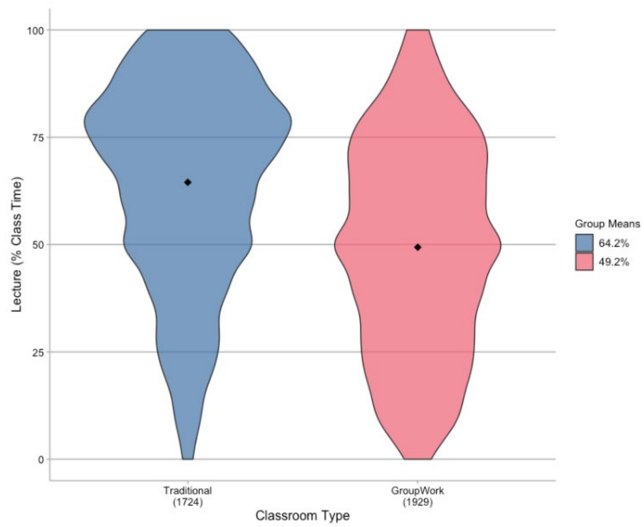

Fig 2. Violin plots of lecture (as percentage of class time) reported by instructors with different classroom types. Group means are indicated in the plots and reported in the legend; N is indicated below each category on the vertical axis; common letters indicate group means that are not statistically different at the 95% confidence level. The amount of lecture used by instructors in different classroom types is different, $t(3612.2) = 18.98$, $p < 0.001$, $g = 0.63$ (medium).

As with class size, it is clear that instructors use active learning in both classroom types, and that there are instructors in both types heavily utilizing lecture. However, and consistent with common belief, there is more lecture in traditional fixed-seat classrooms. The medium effect

[18,19] size of this result indicates a substantive difference in practice between the two classroom types, on average. For change agents, this suggests that having or building classrooms that physically accommodate peer-to-peer interaction is an effective way to support active learning instruction. This is consistent with other research that has found multiple benefits for instructors and institutions for creating active learning classrooms [22]. Another implication of these data is that there are a substantial number of instructors who use little lecture even in traditional fixed-seat classrooms. Thus, it is possible to use active learning in these classrooms and more instructors would likely be able to do so, perhaps with additional support.

Evaluation of teaching

Student evaluations of teaching (SET) are a common, and often contentious, component of the teaching assessments used in review, tenure, and promotion decisions. There are many concerns about the use of SET, including that they are sensitive to gender biases [23] and are not consistently related to learning outcomes [24]. Additionally, we regularly hear that instructors are less likely to use active learning strategies because they fear that use of such strategies may result in lower SET scores [9,10]. We asked two questions about the assessment of teaching: the importance of teaching effectiveness in the overall performance review, and the weighting of SET in the assessment of teaching effectiveness. These five-point Likert-style questions were each collapsed to three levels. Teaching is identified as a small, medium, or large component of performance review and decisions about promotion; SET are given more, equal, or less weight than other measures in the overall assessment of teaching effectiveness.

Among all instructors, an ANOVA test indicates a small effect [18,19] of the importance of teaching evaluation. Instructors who report that teaching assessment is of large importance lecture less than those who report medium or small importance (Fig 3A). To investigate the role of SET, we focused only on those participants who reported that teaching assessment is of large importance. For this subset, ANOVA indicates a small effect of the relative weighting of SET on classroom practice. Those for whom SET is weighted more than other measures (or is the only measure) lecture more than those for whom it is weighted less than other measures (Fig 3B).

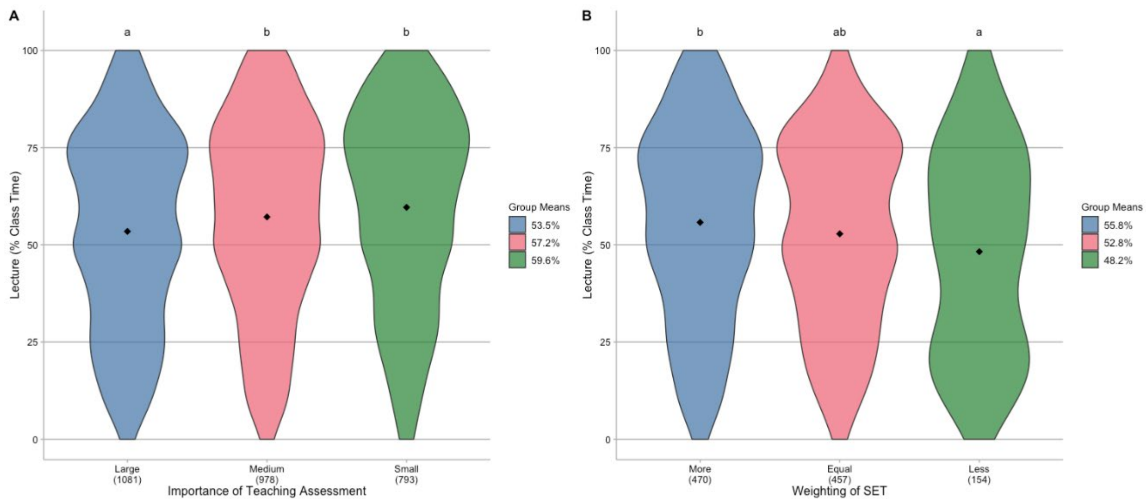

Fig 3. Violin plots of lecture (as percentage of class time) reported by instructors. Group means are indicated in the plots and reported in the legend; N is indicated below each category on the horizontal axis; common letters indicate group means that are not statistically different in Tukey HSD post hoc testing at a 95% confidence level. **A.** All instructors, grouped by the importance of overall teaching assessment for decisions of review and promotion. This is a main effect on lecture, $F(2,2849) = 14.56$, $p < 0.001$, $\eta^2 = 0.01$ (small). **B.** Instructors for whom teaching assessment is of large importance, grouped by the relative weight of SET. This is a main effect on lecture, $F(2,1078) = 5.63$, $p < 0.01$, $\eta^2 = 0.01$ (small).

Thus, our data is consistent with the belief that reliance on SET as the most important part of teaching evaluation impedes the use of active learning. The variation we observe in these data indicate that instructors can use active learning in varied assessment contexts; it also indicates that there are instructors in all assessment contexts heavily utilizing lecture. However, the data show that there is more lecture when teaching assessment is *not* very important in review decisions. Among instructors who report assessment of teaching effectiveness is of large importance in review decisions, lecture increases with increased emphasis on SET (Fig 3b). These results suggest that change agents interested in increasing the use of active learning should work to increase the importance of teaching in performance evaluations and to reduce the importance of SET in the overall evaluation of instruction. This recommendation for reduced emphasis on SET is consistent with research that has found SET are not an appropriate measure of teaching effectiveness and are discriminatory [23,24].

Findings: Individual factors

Security of employment

There are many beliefs about instructional practice in relation to academic rank and experience; these are sometimes contradictory. For example, some suggest that job security (e.g., tenure) allows for the flexibility needed to engage in innovative teaching practice, while others argue that new innovations can only be used by instructors from the newest generation who are more innovative and not yet set in their ways. These beliefs drive practice. For example, many change initiatives focus on future faculty or new instructors [25]. Participants reported whether or not they were on a track leading to increased job security, and we saw no difference in the amount of lecture between these two groups (Fig 4A). For those on secure tracks, we have no evidence of a difference in the amount of lecture used by those who have or have not achieved that security (Fig 4B).

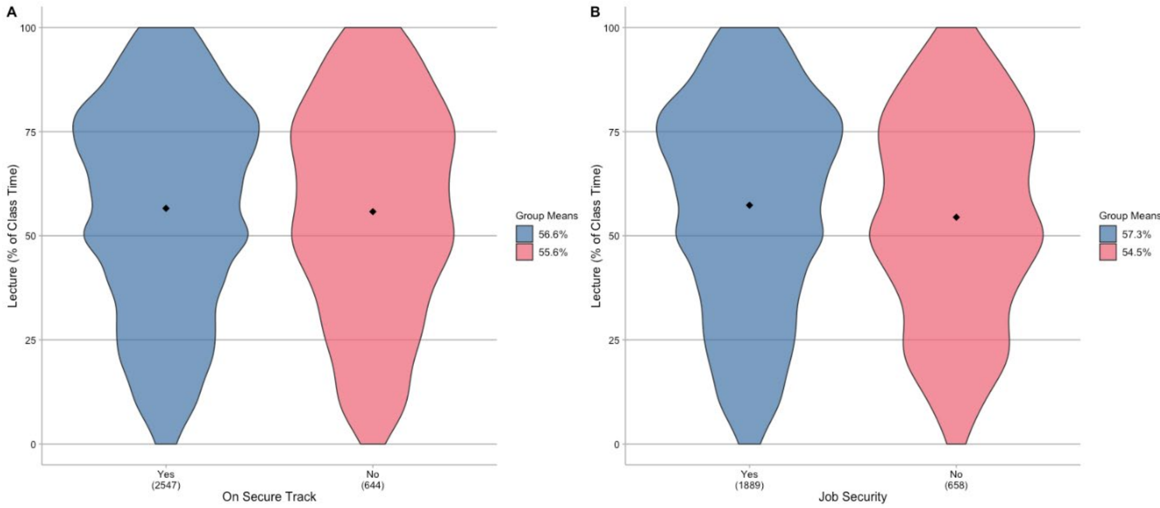

Fig 4. Lecture (as percentage of class time) reported by instructors. Group means are indicated in the plots and reported in the legend; N is indicated below each category on the horizontal axis. **A.** All instructors, grouped by whether or not they are on a track leading to increased job security. There is no evidence of difference between these groups, $t(984.7) = 0.71$, $p > 0.05$, $g = 0.03$ (negligible). **B.** Instructors on a secure track, grouped by whether or not they have achieved increased security. There is no evidence of a difference between these groups, $t(1185.9) = 2.58$, $p < 0.05$, $g = 0.11$ (negligible).

Thus, our survey data does not support beliefs about a relationship between security of employment and the use of active learning. The variation and spread of the data suggest that many instructors, regardless of job security or potential for that security, use active learning strategies - and that in all situations many instructors heavily implement lecture. Change agents seeking to increase the use of active learning should not assume that some instructors are less likely to be receptive to active learning based solely on their security of employment. That said, individuals in more precarious employment situations may be differentially affected by other factors (e.g., if employment is contingent on teaching assessment), which should be taken into consideration when planning professional development or working to change instructional practice.

Research activity

Postsecondary instructors are often expected to balance multiple roles. At many institutions, particularly research-intensive universities, the balance between teaching and research activities is often cited as a barrier to instructional change, with many believing that instructors who focus on research are less innovative in their teaching [9]. We asked respondents about the breakdown of their appointment, including the percentage dedicated to research. First we separated instructors with zero and non-zero research appointments, and found no evidence of a difference in the percentage of class time they report spending on lecture[2].

Among those with a non-zero research appointment, we ranked instructors' research activity level based on self-report of publications, grants, and presentations. There is a small effect [18,19] of research activity level on the percentage of class time spent in lecture. Post hoc comparison of means shows that very active researchers lecture more than others (Fig 5A). These very high researchers reported at least three of the following within the last two years: 20% or greater research appointment; external funding for research; presenting research at two or more professional meetings; submitting two or more manuscripts for publication. Separately, we considered research focus, specifically with an eye toward involvement in education research. Over half of the participants indicated some involvement with education research, which included participating in discipline-based education research, scholarship of teaching and learning, or funded projects aimed at improving undergraduate instruction. There is a medium-sized effect [18,19] of such involvement with education research, and those who *have* been involved use less lecture than those who have not (Fig 5B).

[2] $t(2797) = 1.03$, $p > 0.05$, $g = 0.04$ (negligible)

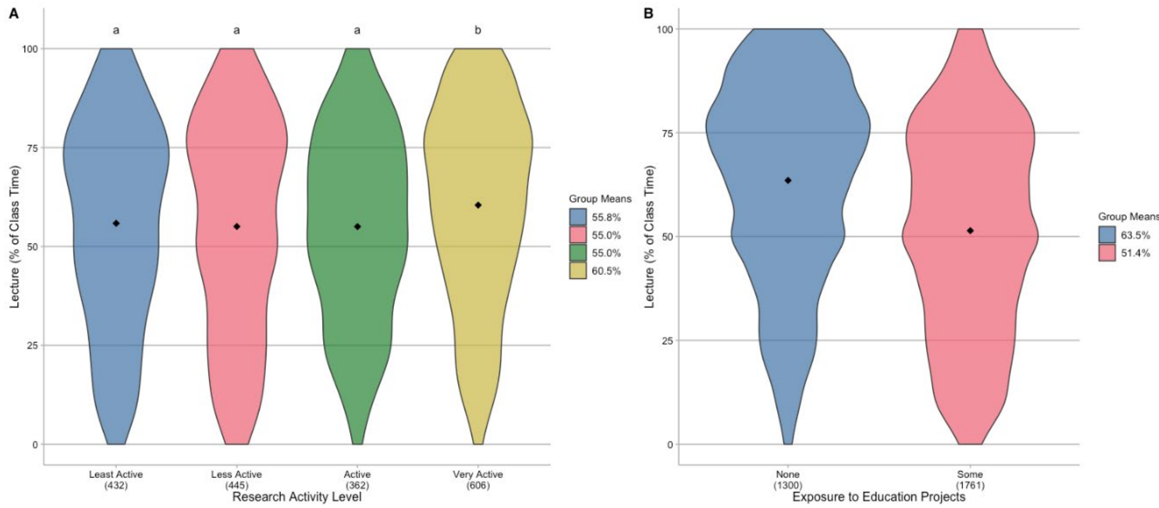

Fig 5. Lecture (as percentage of class time) reported by instructors. Group means are indicated in the plots and reported in the legend; N is indicated below each category on the horizontal axis; common letters indicate group means that are not statistically different in Tukey HSD post hoc testing at a 95% confidence level. **A.** Research activity level is an effect on lecture, $F(3,1840) = 5.60$, $p < 0.001$, $\eta^2 = 0.009$ (small). **B.** Instructors who participate in education research, scholarship of teaching and learning, and/or curricular improvement projects lecture less than those who do not $t(286.2) = 13.6$, $p < 0.001$, $g = 0.49$ (medium).

Thus, our data are at least somewhat consistent with the belief that very high levels of research activity may impede the use of active learning. There are two types of implications for change agents interested in promoting the use of active learning. The first is that the majority of instructors who are engaged in research are not engaged to such an extent that it impacts their use of active learning. The second implication is that those instructors with very high levels of research activity do likely have limited time and will need to be supported in implementing active learning strategies that do not increase, or even decrease, time required [26,27]. Perhaps unsurprisingly, instructors who engage in education research or funded education work are more likely to use active learning. Thus, external and institutional grants for instructional development appear to be valuable strategies for improving instruction. For example, instructors with very high levels of traditional research activity may productively participate in instructional development teams [28,29].

Prior exposure to active learning

It has been suggested that active learning instructional strategies are not likely to be adopted by instructors until they have personal experience with that instructional style [30], and that experiences as a student form instructors' beliefs about teaching [31]. Participants in our survey reported on whether they had any experience as a student in an active learning course or as a student instructor (or instructional team member) in a course taught using active learning. We reduced the sample to only those who responded with "Yes" or "No" to each item; consistent with other work on the uptake of innovative instructional practices in STEM, the majority of instructors in our sample (75%) have not been exposed to such instruction as students. Fig 6 shows that instructors who have experienced active learning as a student report lecturing less than those who have not.

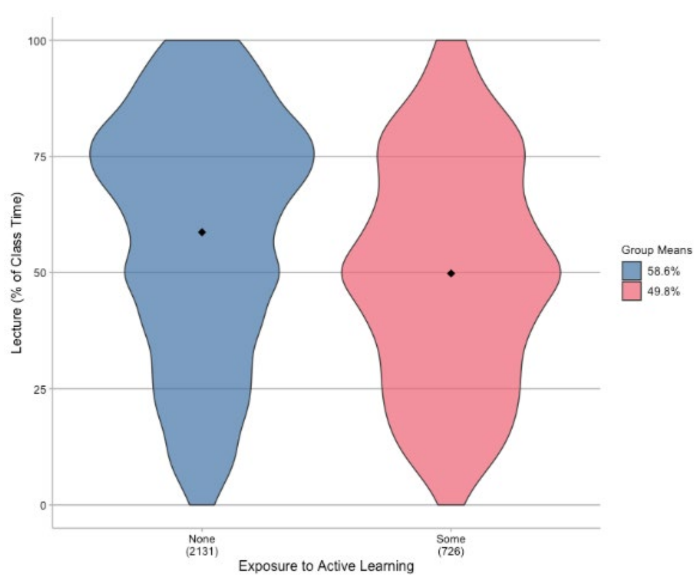

Fig 6. Lecture (as percentage of class time) reported by instructors with different prior experience with active learning. Group means are indicated in the plots and reported in the legend; N is indicated below each category on the horizontal axis. Those who have prior experience with active learning as a student or student instructor lecture less than those who have not $t(1338.7) = 8.46$, $p < 0.001$, $g = 0.35$ (small-medium).

The data support the belief that exposure to active learning increases the likelihood that an instructor will use active learning. For change agents interested in increasing the use of active learning it is, therefore, important to learn about past experiences of instructors and how to build on these. Most of our survey respondents reported having no prior experiences with active learning. It is likely valuable to find ways for them to get such experience, such as co-teaching [32] or participation in a local instructional development team [28,29]. Institutional leaders and change agents should also think about implementation of active learning as not only useful for the current students, but also as the beginning of a cultural change in higher education [9].

Conclusion

We set out to understand the extent to which instructional practice in introductory STEM courses is consistent with six common beliefs about instructor use of active learning instructional strategies. In every context we examined, there are instructors using all active learning and instructors using all lecture. While there is much variation, the patterns in survey data collected from 3769 instructors were consistent with the three beliefs about contextual factors, but varied in consistency with the three beliefs about individual factors:

1. Class size. Our data are *consistent* with the belief that large class sizes can hinder the use of active learning. Instructors of very large classes (100 or more students) report significantly more lecturing than instructors of other classes.
2. Traditional classroom. Our data are *consistent* with the belief that traditional fixed-seat classrooms can hinder the use of active learning. Instructors in classrooms designed for active learning use significantly more active learning.
3. Student evaluations. Our data are *consistent* with the belief that emphasizing student evaluations of teaching can hinder the use of active learning. When assessment of teaching effectiveness is important, instructors at institutions with less emphasis on student evaluations report more active learning.
4. Security of employment. Our data are *not consistent* with the belief that not having security of employment (e.g., tenure) can hinder the use of active learning. Instructors without security of employment report using similar levels of active learning to those with security.
5. Research activity. Our data are *somewhat consistent* with the belief that high levels of research activity can hinder the use of active learning. Instructors with very high research productivity report using active learning less than other instructors. On the other hand, instructors who engage in education research or funded curriculum development use more active learning.
6. Experience as a student. Our data are *consistent* with the belief that exposure to active learning as a student or student instructor supports the use of active learning. Instructors who were students in an active learning classroom and/or were part of an instructional team that used active learning are more likely to use active learning instruction.

Those in positions to make structural or policy changes should note these findings and incorporate them into decision-making processes. Institutions should seek to make targeted structural changes that support active learning instruction, such as maintaining and seeking smaller class sizes, building and supporting active learning classrooms, and emphasizing methods beyond SET scores for the evaluation of teaching effectiveness.

These results also suggest the importance of professional development opportunities, both local and national. As the results show, instructors with all types of individual characteristics and in all types of contexts do manage to incorporate large amounts of active learning into their instruction. Some forms of active learning instruction are more compatible with large class sizes than others. Instructors are more likely to implement active learning if they do not have to figure out everything on their own. Professional development activities that focus on local needs and contexts can be highly effective. This is consistent with many of the current department-based change initiatives [28,29,33].

Our findings reflect national trends in the use of active learning by instructors with different contextual and individual characteristics. There are, of course, many individual instructors who buck these trends and incorporate active learning in their courses regardless of physical setups and cultural norms which might be further unpacked by future research. The variation we observe in our data suggests that active learning is possible by any instructor in any environment; however, policies can and should be enacted to facilitate and support individual choices to use more active learning in undergraduate STEM courses.

Acknowledgements

We would like to acknowledge and thank the participants in our study for their time and effort telling us about their instructional practices and context. Similarly, the members of the American Institute of Physics Statistical Research Center for their help gathering the survey data for this study. Additional thanks to Dana Kirin for consulting with the first author about the data analysis.

References

1. Koch AK. It's About the gateway courses: Defining and contextualizing the issue. New Directions for Higher Education. 2017;2017: 11–17. doi:10.1002/he.20257
2. Seymour E, Hewitt NM. Talking About Leaving: Why Undergraduates Leave the Sciences. Boulder, CO: Westview Press; 1997.
3. Seymour E, Hunter A-B. Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education. 2019; doi:10.1007/978-3-030-25304-2
4. NRC (National Research Council). Transforming undergraduate education in science, mathematics, engineering, and technology. Washington, DC: National Research Council; National Academy Press; 1999.
5. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences. 2014;111: 8410–8415. doi:10.1073/pnas.1319030111

6. Theobald EJ, Hill MJ, Tran E, Agrawal S, Arroyo EN, Behling S, et al. Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. Proc Natl Acad Sci USA. 2020; 201916903. doi:10.1073/pnas.1916903117
7. Rasmussen C, Apkarian N, Hagman JE, Johnson E, Larsen S, Bressoud D, et al. Characteristics of Precalculus through Calculus 2 programs: Insights from a national census survey. Journal for Research in Mathematics Education. 2019;50: 98–112.
8. Stains M, Harshman J, Barker MK, Chasteen SV, Cole R, DeChenne-Peters SE, et al. Anatomy of STEM teaching in North American universities. Science. 2018;359: 1468–1470. doi:10.1126/science.aap8892
9. Shadle SE, Marker A, Earl B. Faculty drivers and barriers: Laying the groundwork for undergraduate STEM education reform in academic departments. International Journal of STEM Education. 2017;4. doi:10.1186/s40594-017-0062-7
10. Sturtevant H, Wheeler L. The STEM Faculty Instructional Barriers and Identity Survey (FIBIS): Development and exploratory results. IJ STEM Ed. 2019;6: 35. doi:10.1186/s40594-019-0185-0
11. Gibbons RE, Villafañe SM, Stains M, Murphy KL, Raker JR. Beliefs about learning and enacted instructional practices: An investigation in postsecondary chemistry education. J Res Sci Teach. 2018;55: 1111–1133. doi:10.1002/tea.21444
12. Lund TJ, Stains M. The importance of context: An exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. IJ STEM Ed. 2015;2: 13. doi:10.1186/s40594-015-0026-8

13. Apkarian N, Smith WM, Vroom K, Voigt M, Gehrtz J, PtC Project Team, et al. X-PIPS-M Survey Suite. 2019. Available: https://www.maa.org/sites/default/files/XPIPSM%20Summary%20Document.pdf
14. Johnson E, Keller R, Fukawa-Connelly T. Results from a survey of Abstract Algebra instructors across the United States: Understanding the choice to (not) lecture. International Journal of Research in Undergraduate Mathematics Education. 2018;4: 254–285. doi:10.1007/s40753-017-0058-1
15. Henderson C, Dancy MH. Impact of physics education research on the teaching of introductory quantitative physics in the United States. Phys Rev ST Phys Educ Res. 2009;5: 020107. doi:10.1103/PhysRevSTPER.5.020107
16. Walter EM, Henderson CR, Beach AL, Williams CT. Introducing the Postsecondary Instructional Practices Survey (PIPS): A concise, interdisciplinary, and easy-to-score survey. Allen D, editor. CBE—Life Sciences Education. 2016;15: ar53. doi:10.1187/cbe.15-09-0193
17. Durham MF, Knight JK, Bremers EK, DeFreece JD, Paine AR, Couch BA. Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). IJ STEM Ed. 2018;5: 31. doi:10.1186/s40594-018-0128-1
18. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates; 1988; doi:10.4324/9780203771587
19. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013;4. doi:10.3389/fpsyg.2013.00863

20. Fagen AP, Crouch CH, Mazur E. Peer Instruction: Results from a range of classrooms. The Physics Teacher. 2002;40: 206–209. doi:10.1119/1.1474140

21. Mazur E. Peer instruction: A user's manual. Upper Saddle River, N.J: Prentice Hall; 1997.

22. Knaub AV, Foote KT, Henderson C, Dancy M, Beichner RJ. Get a room: The role of classroom space in sustained implementation of studio style instruction. IJ STEM Ed. 2016;3: 8. doi:10.1186/s40594-016-0042-3

23. Mengel F, Sauermann J, Zölitz U. Gender Bias in Teaching Evaluations. Journal of the European Economic Association. 2019;17: 535–566. doi:10.1093/jeea/jvx057

24. Uttl B, White CA, Gonzalez DW. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. Studies in Educational Evaluation. 2017;54: 22–42. doi:10.1016/j.stueduc.2016.08.007

25. CSSP. The role of scientific societies in STEM faculty workshops. Washington, DC: Council of Scientific Society Presidents; 2013. Available: https://www.aapt.org/Conferences/newfaculty/upload/STEM_REPORT-2.pdf

26. Gafney L, Varma-Nelson P. Peer-Led Team Learning: Evaluation, Dissemination, and Institutionalization of a College Level Initiative. Dordrecht: Springer Netherlands; 2008. doi:10.1007/978-1-4020-6186-8

27. Tien LT, Roth V, Kampmeier JA. Implementation of a peer-led team learning instructional approach in an undergraduate organic chemistry course. J Res Sci Teach. 2002;39: 606–632. doi:10.1002/tea.10038

28. Olmstead A, Beach A, Henderson C. Supporting improvements to undergraduate STEM instruction: an emerging model for understanding instructional change teams. IJ STEM Ed. 2019;6: 20. doi:10.1186/s40594-019-0173-4

29. Reinholz DL, Corbo JC, Dancy M, Finkelstein N. Departmental action teams: Supporting faculty learning through departmental change. Learning Communities Journal. 2017;9: 5–32.

30. Andrews TC, Lemons PP. It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. Tomanek D, editor. LSE. 2015;14: ar7. doi:10.1187/cbe.14-05-0084

31. Oleson A, Hora MT. Teaching the way they were taught? Revisiting the sources of teaching knowledge and the role of prior experience in shaping faculty teaching practices. High Educ. 2014;68: 29–45. doi:10.1007/s10734-013-9678-9

32. Henderson C, Beach A, Famiano M. Promoting instructional change via co-teaching. American Journal of Physics. 2009;77: 274–283. doi:10.1119/1.3033744

33. Soto RC, Marzocchi AS. Learning about Active Learning While Actively Learning: Insights from Faculty Professional Development. PRIMUS. 2020; 1–12. doi:10.1080/10511970.2020.1746449

34. RStudio Team. RStudio: Integrated development environment for R. Boston, MA: RStudio, Inc.; 2020. Available: http://www.rstudio.com

35. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available: https://www.R-project.org

36. Wickham H, Bryan J. readxl: Read Excel Files. 2019. Available: https://CRAN.R-project.org/package=readxl

37. Wickham H. Reshaping Data with the reshape Package. J Stat Soft. 2007;21. doi:10.18637/jss.v021.i12

38. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, IL: Northwestern University; 2019. Available: https://CRAN.R-project.org/package=psych

39. Torchiano M. effsize: Efficient Effect Size Computation. Zenodo; 2016. doi:10.5281/ZENODO.1480624

40. Lüdecke D. sjstats: Statistical Functions For Regression Models. Zenodo; 2018. doi:10.5281/ZENODO.1284472

41. Hothorn T, Bretz F, Westfall P. Simultaneous Inference in General Parametric Models. Biom J. 2008;50: 346–363. doi:10.1002/bimj.200810425

42. Wickham H. ggplot2: Elegant graphics for data analysis. Second edition. Cham: Springer; 2016.

43. Kassambara A. ggpubr: “ggplot2” based publication ready plots. 2020. Available: https://CRAN.R-project.org/package=ggpubr

44. Arnold JB. ggthemes: Extra themes, scales, and geoms for “ggplot2.” 2019. Available: https://CRAN.R-project.org/package=ggthemes

Supporting information

Methods details
Tables S1-S5
References 24-39

Methods details

Data cleaning and reduction

1. During a typical week, what proportion of time during regular class meetings (i.e., lecture sections) do students spend doing the following?
    a. Working individually [Dropdown: 0, 5, 10, …, 95, 100]%
    b. Working in small groups [Dropdown: 0, 5, 10, …, 95, 100]%
    c. Participating in whole class discussion [Dropdown: 0, 5, 10, …, 95, 100]%
    d. Listening to the instructor lecture or solve problems [Dropdown: 0, 5, 10, …, 95, 100]%

Participants were asked to make sure their selections totaled 100. We received 3641 responses which totaled to 100%, 128 which totaled to 0%, and 26 which totaled to something else. The 154 which did not total 100 were removed from the analysis.

2. What is the approximate enrollment in a typical lecture section?
    a. [Text entry]

This resulted in a wide range of entry formats, including ranges (e.g., 20-30) as well as specific values. As we do not expect that people make instructional decisions based on the exact number of students enrolled (e.g., 24 vs. 26 students), we elected to bin enrollments, and to do so based on the average of ranges when those were provided. Three factors were considered in creating these bins: (1) the distribution of responses; (2) our own knowledge of active learning design and usage; and (3) the well-known bins selected by the US News & World report rankings (https://www.usnews.com/education/best-colleges/articles/ranking-criteria-and-weights). These bins are: 0-19, 20-29, 30-39, 40-59, 60-99, 100+. In total, 3725 participants answered this item. Four were removed because they answered a different question (e.g., "50 minutes") and 65 provided ranges, which were then averaged to determine the appropriate bin.

3. Which of the following best describes the set-up in your classroom?
    a. Classroom with fixed seats
    b. Classroom that accommodates group work
    c. Other [Text entry option]

Participants found this question challenging to answer, evidenced by the 228 "other" responses which were submitted along with some form of description or not. These were reviewed by the project team, and a decision was made to adjudicate responses into two groups: (1) physical classrooms that *easily accommodate* group work; (2) physical classrooms that *do not easily accommodate* group work. The original limitation did not account for a range of reported rooms, such as those with fixed rows of tables but with movable chairs.

4. What is the role of student evaluations of teaching (SET) in evaluating teaching performance in decisions of review, promotion, or tenure?
    a. SET are the only measure used to evaluate teaching performance
    b. SET are used and given more weight as compared to other measures
    c. SET are used and given equal weight as compared to other measures

    d. SET are used and given less weight as compared to other measures
    e. SET are not used to evaluate teaching performance
5. How much does the overall assessment of teaching performance matter in decisions of review, promotion, or tenure for someone in your role?
    a. It is not considered
    b. Somewhat influential
    c. Influential
    d. Very influential

Based on the distribution of responses, each of these scales was collapsed to three levels. For #4, (a) and (b) were merged, as were (d) and (e). For #5, (a) and (b) were merged.

6. What is your present academic rank?
    a. Professor
    b. Associate Professor
    c. Assistant Professor
    d. Lecturer/Instructor
    e. Visiting Professor/Lecturer/Instructor
    f. Postdoctoral Instructor
    g. Graduate Student Instructor or Teaching Assistant
7. What is your tenure status at this institution?
    a. Tenured
    b. On tenure track, but not tenured
    c. Not on tenure track, but this institution has a tenure system
    d. No tenure system at this institution
8. Do you have the opportunity for promotion that comes with increased security of employment?
    a. Yes, and I <u>have</u> received such a promotion
    b. Yes, and I <u>have not</u> received such a promotion
    c. No

Those who selected a-d in question #6 were shown question #7; those who selected c-d in question #7 were shown question #8. Classification of employment with the possibility of promotion is associated with 7ab and 8ab; having achieved additional security is associated with 7a and 8a.

9. What is the approximate distribution of your position at [*institution name*]?
    a. Research [Dropdown: 0, 5, 10, …, 95, 100]%
    b. Teaching [Dropdown: 0, 5, 10, …, 95, 100]%
    c. Service [Dropdown: 0, 5, 10, …, 95, 100]%
    d. Administration [Dropdown: 0, 5, 10, …, 95, 100]%
    e. Other [Dropdown: 0, 5, 10, …, 95, 100]%
10. Do you currently have external funding for research?
    a. Yes
    b. No

11. How many professional meetings/conferences have you presented your research or scholarship within the past two years?
    a. [Drop down: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+]
12. How many articles about your research or scholarship have you submitted for publication within the past two years?
    a. [Drop down: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10+]

For #1, participants were asked to ensure that their responses totaled 100, but of the responses, 554 totaled zero (equivalent to skipping the item) and 49 totaled something else. These were removed, leaving 3166 responses which totaled 100 for analysis.

Median responses for each of these four items were calculated, and respondents received a "research activity point" for each if they were above the median, corresponding to meeting the following criteria:

- Research appointment > 20% (#9)
- External research funding (#10)
- Presenting at two or more professional meetings (#11)
- Submitting two or more articles for publication (#12)

These scores were then converted to activity levels:

- 0 – Least active
- 1 – Less active
- 2 – Active
- 3-4 – Very active

13. Do you conduct STEM education research and/or participate in the scholarship of teaching and learning?
    a. Yes
    b. No
    c. I don't know
14. In the last five years, have you been part of a project that has received any of the following?
    a. National Science Foundation funding to improve an undergraduate course or the undergraduate curriculum [Yes/No]
    b. External funding (not from the National Science Foundation) to improve an undergraduate course or the undergraduate curriculum [Yes/No]
    c. Internal funding (i.e., from your institution) to improve an undergraduate course or the undergraduate curriculum [Yes/No]

Response to questions #13 and #14 were combined to create an indicator of "involvement with education improvement research." Those who provided a single "yes" response to any of #13 or #14a-c are considered to have some involvement, while those who provided a "no" to each are not. Responses of "I don't know" in Question #13 were treated as having skipped the item rather than lumped with either "yes" or "no" responses. Analysis was conducted on the data of the 3061 participants who answered both items.

15. Have you ever been a student in a course taught using RBIS?
    a. Yes
    b. No
    c. I don't know
16. While a student, have you ever been part of an instructional team for a course taught using RBIS?
    a. Yes
    b. No
    c. I don't know

Participants who responded with "I don't know" were removed from the analysis, and a yes to either (or both) questions was considered evidence of prior exposure to active learning while a student.

Data analysis
Data reduction, cleaning, combining, and analyses were conducted using RStudio version 1.2.5042 [34] and R version 3.6.2 [35]. Packages used for data management, analysis, and development of figures: readxl [36], reshape2 [37], psych [38], effsize [39], sjstats [40], multcomp [41], ggplot2 [42], ggpubr [43], ggthemes [44].

When our comparisons involve exactly two groups, we use Welch's unequal variance *t*-test. In addition to statistical significance, we report standardized effect sizes using Hedge's *g*. When our comparisons involve factors with more than two levels, we use ANOVA. In addition to reporting *F*-statistics and *p*-values, we report standardized η^2 estimators of variance. Post-hoc testing is conducted with Tukey's HSD test, with a family-wise 95% confidence level.

Additional details of results

Table S1. Summary statistics: percentage of class time spent in lecture by target groups.

	Count	Mean	SD	Median	Minimum	Maximum
Binned class enrollment[1]						
0-19	380	52.49	24.34	50	0	100
20-29	973	52.32	24.12	50	0	100
30-39	740	55.47	25.26	60	0	100
40-59	587	55.94	25.36	60	0	100
60-99	378	59.43	25.86	65	0	100
100+	578	66.54	24.02	70	0	100
Classroom setup[1]						
Traditional	1717	64.50	23.68	70	0	100
Accommodates group work	1923	49.38	24.38	50	0	100
Importance of teaching assessment for review[1]						
Large	1081	53.46	24.85	55	0	100
Medium	978	57.17	25.11	60	0	100
Small	793	59.64	25.26	65	0	100
Weighting of SET compared to other measures of teaching assessment[2]						
Heavy / Only	470	55.78	24.95	60	0	100
Equal	457	52.82	23.96	50	0	100
Light / none	154	48.25	26.31	50	0	100
On secure track[1]						
Yes	2547	56.58	25.20	60	0	100
No	642	55.79	25.35	60	0	100
With increased security (e.g., tenure)[3]						
Yes	1889	57.33	25.42	60	0	100
No	658	54.45	24.46	55	0	100
Percentage of appointment dedicated to research[1]						
Zero	1269	55.99	24.51	60	0	100
Non-zero	1905	56.87	25.59	60	0	100
Research Activity Level[4]						
Inactive	431	55.85	25.66	60	0	100
Less active	445	55.04	27.19	60	0	100
Active	362	55.01	23.83	57.5	0	100
Very active	606	60.45	25.08	60	0	100
Involvement with DBER, SoTL, and/or funded education project[1]						
None	1299	63.53	23.94	70	0	100
Some	1761	51.42	24.99	50	0	100
Experience with active learning as a student and/or student instructor[1]						
No	2127	58.68	25.71	60	0	100
Yes	726	49.81	23.92	50	0	100

1. All participants who answered the question(s).
2. All participants who answered the question and reported that teaching assessment is very important in decisions of review.
3. All participants who answered the question and reported that they are on a track with the possibility of increased security.
4. All participants who answered the question and reported a non-zero research appointment.

Table S2. Results of Welch two sample t-tests: percentage of class time spent in lecture by target groups.

	t	df	Sig	Mean diff.	95% CI of difference		Hedges' g	95% CI of g		Size of g
					Lwr.	Uppr.		Lwr.	Uppr.	
Classroom setup[1]	18.98	3612.2	<0.001	15.12	13.56	16.69	0.63	0.56	0.70	Medium
Secure track[1]	0.71	984.7	ns	0.79	-1.39	2.99	0.03	-0.05	0.12	Trivial
Increased security[2]	2.58	1185.9	0.010	2.88	0.69	5.08	0.11	0.03	0.20	Trivial
Research appointment[1]	0.96	2796.5	ns	0.87	-0.90	2.65	0.03	-0.04	0.11	Trivial
DBER involvement[1]	13.57	286.2	<0.001	12.11	10.36	13.86	0.49	0.42	0.57	Small-Medium
Student experience[1]	8.46	1338.7	<0.001	8.87	6.81	10.93	0.35	0.27	0.44	Small

1. All participants who answered the question(s).
2. All participants who answered the question and reported that they are on a track with the possibility of increased security.

Table S3. Analysis of variance models: percentage of class time spent in lecture by target groups.

	Statistic	Sig.	η^2	95% CI of η^2		Size of η^2
				Lower	Upper	
Class size[1]	$F(5, 3630) = 27.9$	< 0.001	0.037	0.025	0.049	Small
Importance of teaching assessment[1]	$F(2, 2849) = 14.56$	< 0.001	0.010	0.003	0.018	Small
Weight of SET in teaching assessment[2]	$F(2, 1078) = 5.63$	0.004	0.010	-0.001	0.025	Small
Research activity level[3]	$F(3, 1840) = 5.60$	< 0.001	0.009	0.002	0.018	Small

1. All participants who answered the question(s).
2. All participants who answered the question and reported that teaching assessment is very important in decisions of review.
3. All participants who answered the question and reported a nonzero research appointment.

Table S4. Tukey HSD test at 95% family-wise confidence level: percentage of class time spent in lecture by target groups.

Group 1	Group 2	Mean diff.	Lower	Upper	*p* (adjusted)
Course enrollment					
20-29	0-19	-0.16	-4.43	4.11	ns
30-39	0-19	2.99	-1.47	7.44	ns
40-59	0-19	3.45	-1.20	8.10	ns
60-99	0-19	6.94	1.82	12.07	0.002
100+	0-19	14.05	9.39	18.71	< 0.001
30-39	20-29	3.15	-0.29	6.60	ns
40-59	20-29	3.61	-0.07	7.30	ns
60-99	20-29	7.11	2.83	11.39	< 0.001
100+	20-29	14.22	10.51	17.92	< 0.001
40-59	30-39	0.46	-3.44	4.36	ns
60-99	30-39	3.96	-0.50	8.42	ns
40-59	30-39	11.07	7.15	14.98	< 0.001
60-99	40-59	3.49	-1.16	8.15	ns
100+	40-59	10.60	6.47	14.74	< 0.001
100+	60-99	7.11	2.44	11.78	< 0.001
Importance of teaching assessment					
Medium	Big	3.72	1.13	6.31	0.002
Small	Big	6.19	3.44	8.93	< 0.001
Small	Medium	2.47	-0.34	5.28	ns
Weight of SET in teaching assessment					
Equal	Heavy	-2.95	-6.77	0.86	ns
Light	Heavy	-7.53	-12.92	-2.14	< 0.001
Light	Equal	-4.58	-9.99	0.83	ns
Research activity level					
Less	Least	-0.80	-5.23	3.63	ns
Active	Least	-0.83	-5.51	3.84	ns
Very	Least	4.61	0.47	8.74	0.022
Active	Less	-0.03	-4.67	4.61	ns
Very	Less	5.41	1.31	9.50	0.004
Very	Active	5.44	1.08	9.80	0.007