Interactive problem-solving sessions in an introductory bioscience course engaged students and gave them feedback, but did not increase their exam scores

James P. McEvoy\*

**\*Corresponding author:** *Centre for Biomedical Sciences, School of Biological Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK. Tel: (+44 1784 443770); E-mail:* [*james.mcevoy@rhul.ac.uk*](mailto:james.mcevoy@rhul.ac.uk)

**One sentence summary:** In-class problem-solving sessions using Pearson’s Learning Catalytics gave students feedback and boosted their perceived engagement, but did not improve their test scores.

**ABSTRACT**

Active learning, including the promotion of student interactivity in lectures, has been found to improve student engagement and performance in university science classes. This letter describes the use of Pearson’s Learning Catalytics to run regular, formatively-assessed problem-solving sessions as part of the semi-flipped redesign of an introductory level university bioscience course. Students found the problem solving sessions more engaging than a traditional lecture, and felt that they were receiving better feedback on their progress in the course. Their participation in the problem solving sessions was strongly associated with their performance in the course’s summative assessments, making it possible to identify and assist probable poor performers early in the course. Other measures of student engagement with the course were not improved, and neither were their average exam grades when compared with their grades in a course which had not been redesigned. Possible reasons for this are discussed.

**INTRODUCTION**

Active learning can be broadly defined as anything students might be asked to do in class besides listening to a teacher and taking notes (Felder & Brent, 2016). A class or lecture that makes use of active learning may be described as interactive. Interactivity in science classes often involves students solving problems, generally in groups (McClean & Crowe, 2017, Stockwell *et al.*, 2017), and there is mounting evidence that this kind of activity improves the engagement and performance of university students compared to a traditional lecture course (Michael, 2006, Preszler, 2009, Haak *et al.*, 2011, Freeman *et al.*, 2014).

This letter describes a multi-year comparison of traditional lecturing to regular, interactive problem-solving in a large first year undergraduate (level 4) bioscience course at a UK university. Learning Catalytics (Pearson, 2017), a web-based student response system based on the work of Eric Mazur and co-workers (Julie *et al.*, 2013) and used in several other recent studies (Mooring *et al.*, 2016, Rinaldi *et al.*, 2016, Stockwell *et al.*, 2017), was used to run formatively-assessed problem-solving sessions for half of the time previously devoted to lecturing. The displaced course content was taught to students through online lectures, constituting a blended, semi-flipped approach (Eichler & Peeples, 2016). Student performance was summatively assessed by a final written exam (70%) and by various pieces of coursework (30%). Engagement and satisfaction with the course were measured by a questionnaire at the end of the course. It was hypothesized that the interactive teaching method would 1) provide a means of predicting student success in the final exam, to allow timely intervention; 2) increase student engagement with the course; and 3) boost overall student performance in the course’s final written exam.

**STUDY DESIGN**

The study centred on *Principles of Molecular Bioscience* (BS1030)*,* a 30-credit introductory undergraduate course in the School of Biological Sciences at Royal Holloway, University of London. It is a biological chemistry module that is mandatory for students enrolled in various degree programmes, including BSc Biomedical Sciences and BSc Biochemistry. Student data from BS1030 was compared with other years’ data from the same course and where possible with that obtained from the same students in BS1090, the control course in this study. BS1090 is a 30-credit biochemistry course that follows immediately after BS1030, in Term 2 of Year 1, is mandatory for the same set of degree programmes, and remained predominantly lecture-based over the period of the study. An average of 111 students completed BS1030 each year over the four academic years of the study, from 2013-14 to 2016-17. During this period a total of 425 students completed both BS1030 and BS1090, with 197 students taking both courses before the BS1030 redesign and 228 students after.

BS1030 ran in Term 1 each year (late September to early December), and included approximately 33 hours of teaching in the lecture theatre (two 100-minute classroom sessions per week, for 10 weeks) and the same time in the laboratory (one 5-hour laboratory period per week). A formative mid-term test was adminstered halfway through the course. Each classroom session comprised two 50-minute teaching periods separated by a 10-minute break. Written lab reports counted for 10% of the final grade, and a practical exam (introduced in 2015) also contributed 10%. Multiple choice question (MCQ) quizzes were administered online before each lecture and lab session, which together contributed 10% to the student’s final grade. Student performance in BS1030 was assessed summatively primarily through a 2-hour written exam in January (70% of final grade), which comprised two equally-weighted sections: an MCQ section and a short-answer section. This author was the teacher and designer of all aspects of the course.

The study design is summarised in Scheme 1. Before the adoption of the more interactive teaching method (*i.e.* 2013-14 and 2014-15) students were asked to prepare for each lecture by reading the assigned section of the textbook (Crowe & Bradshaw, 2014) and answering an online quiz on this material. They would then attend a classroom session comprising two 50-minute lectures separated by a 10-minute break and punctuated by one or two clicker questions (TurningPoint, Turning Technologies). In 2015-16 and 2016-17 students were instead asked to prepare for each of the eighteen classroom teaching sessions by both watching an online lecture (a screencast) and reading the assigned section of the textbook, and then answering an online MCQ quiz based on both sources. Each classroom session comprised one 50-minute “live” lecture followed by a 10-minute break and a subsequent 50-minute problem-solving session using Learning Catalytics. Audio recordings of the live lectures (but not the problem-solving sessions) were made with Panopto and were available to students after the sessions, along with synchronized Microsoft PowerPoint slides.

Students recieved and responded to Learning Catalytics questions on their own web-enabled device, usually a smartphone, while signed in to their own account. Two departmental tablet computers were available for students to borrow if they needed to. Various question types were used, including composite sketch, direction, image upload, many choice, matching, multiple choice, numerical, ranking, region, short answer and word cloud. Students were given several minutes to discuss each question with their neighbour and, in 2016-17, a postgraduate teaching assistant, and then asked to answer individually. After all the students had been given a chance to answer the question, the anonymized class results were shown to the students through their devices and on the projector screen. The question types amenable to automatic grading (direction, many choice, matching, multiple choice, numerical, ranking and region) were marked by the system and the score recorded in the Learning Catalytics gradebook. Generic feedback was issued and recorded, including the correct answer where appropriate. This was immediately followed by a class discussion on the question, led by this author. Students signalled the extent of their understanding in real time using the “I get it now” or “I still don’t get it” buttons on their screens. All of the questions, their own answers and the anonymized class data were available to each student until the end of the course. All of the data, including student names, were available to this author as an instructor on the course. An example Learning Catalytics question of the “region” type is shown in Figure 1, along with the anonymized class answers.

After the final revision session for the course, Royal Holloway University course evaluation forms were given in hard copy to those students in attendance. These contained 24 questions on various aspects of the course, graded on a Likert scale, as well as a free-response text section. The Likert scale questions were organized in four sections: “About the Course”, “About your Engagement with the Course”, “About the Lecturing Staff for the Course”, and “About the Practical Sessions”. Comparable evaluation data from 2013-14 were not available for this study because the College’s evaluation forms were redesigned after that academic year.

**RESULTS**

The number of Learning Catalytics sessions in which a student participated, by logging in and attempting at least one question, was used as a simple measure of a student’s involvement in the problem-solving sessions. Over the two post-intervention years of the study the median number of sessions in which each student participated was 15 out of the 18 available each year. The number of Learning Catalyticslogins was found to be positively correlated with success in the course above about 7 logins (Figure 2). This correlation with success in the summative assessments was stronger than with any other readily available measure, including lecture attendance and the time a student spent watching the screencasts (data not shown). Our first hypothesis – that running problem-solving sessions with Learning Catalytics would provide a contemporaneous way of predicting a student’s success in the course – was therefore confirmed.

Our second hypothesis was that running problem-solving sessions with Learning Catalytics would boost students’ engagement with the course. Because the course evaluation questionnaires used to collect engagement data were anonymous, it was not possible to measure the relationship between a student’s use of Learning Catalytics and their self-reported engagement. Instead, we made a static group comparison of average self-reported engagement before the intervention (2014-15) with that obtained after the intervention (2015-16 and 2016-17).

There was no consistent, significant improvement in the section of the questionnaire entitled “About your Engagement with the Course”. This included the statements “I worked on this course beyond attending the lectures,” “I attended lectures regularly” and “I was well prepared for lectures”. There was, however, a consistent and statistically significant improvement (two-way chi-square test, P < 0.05) in the responses to two statements in other sections of the questionnaire. One of these was “The teaching style was engaging”, and the other was “I received helpful feedback on my progress during the course”. Figure 3 shows that the extent to which students agreed with both of these statements increased from 2014-15 to 2015-16, and remained more or less unchanged in 2016-17. This change was not mirrored in BS1090, the control course. Therefore our second hypothesis was supported: although the institution’s primary measures of student engagement were not affected, the interactive teaching style was apparently perceived as more engaging than a traditional lecture, and students also felt that they were receiving more helpful feedback on their progress.

Our third hypothesis, that average student performance in the final written exam would be improved by the introduction of the regular problem-solving sessions, was tested by comparing the BS1030 exam results before and after the change in teaching method with the same students’ results in BS1090. Figure 4 shows that although the median mark for the BS1030 exam showed a slight and uneven upward trend over the four academic years, there is no evidence to suggest that the BS1030 exam performance was significantly improved relative to BS1090 by the introduction of a more interactive teaching method in 2015-16. The reduction in the spread of marks after the intervention, to the right of the dashed line in Figure 4, was seen in both BS1030 and BS1090, and corresponded in both courses to a reduction in failure rate (overall course grade <40%). The increase in marks in 2016-17, which was probably caused by a change in MCQ marking method within the department, was greater in BS1090 than in BS1030. Our third hypothesis was therefore not supported by our data.

**DISCUSSION**

Active learning is more popular with educationalists than it is with university science teachers. One reason for this is that student-centred learning was originally promoted in a way which tended to depreciate the transmission of knowledge (see, for example, (Phillips, 1982)). Such constructivism is regarded with scepticism by many scientists, who see themselves as dealing with an objective reality that can successfully be conveyed to students (Taber, 2016). This viewpoint favours the traditional lecture and tends to be suspicious of techniques that reduce the transmission of factual content. In recent years educationalists have largely abandoned psychological justifications for their pedagogies and attempted instead to provide evidence for the effectiveness of active learning in the form of improved test scores and student engagement. At the same time, technology has made it possible to adopt active learning techniques without sacrificing course content (Eichler & Peeples, 2016). The present work adds to the recent empirical literature on the effectiveness of active learning techniques as part of a partially flipped course design.

We found that the number of formative, in-class problem-solving sessions in which a student participated correlated with the student’s performance in the course’s summative assessments (Figure 2), of which the final written exam was the main component. This result suggests that participation in such problem-solving sessions may be used as a contemporaneous proxy for a student’s engagement with a course, and provide a useful predictor of performance. The 29 students who participated in seven or fewer problem-solving sessions obtained an average grade of less than 48% in the course overall, and six of them failed the course. In contrast, the 44 students who participated in all eighteen of the Learning Catalytics sessions averaged 64% in the course overall, and none of them failed. This result suggests that it is possible to use Learning Catalytics participation data quite early in the course to identify students at high risk of failure and to direct them towards suitable learning support.

The students’ answers to the course evaluation questionnaires suggested that they were more engaged by the interactive teaching design than by traditional lectures alone, and appreciated the increased feedback on their progress that they received. The improvement in perceived feedback was unexpected, and welcome. It presumably came about because students felt the problem-solving sessions provided a regular test of their understanding, followed immediately by a class discussion which provided high-quality (although necessarily generic) verbal feedback. This result is a reminder that useful feedback to students need not be in the form of time-consuming written or recorded comments, but may instead be given as part of regular interactive teaching sessions within scheduled class time.

The change in teaching method did not improve the results in the section of the questionnaire designed by this author’s institution to measure students’ overall engagement with the course. This section measured the extent to which students worked on the course beyond attending lectures, the students’ self-reported attendance at lectures, and how well students felt they were prepared for lectures. It is remarkable that asking students to watch an online lecture and answer an associated summative quiz before every live teaching session (Scheme 1) seemed to increase neither the time that students spent working on the course, nor the extent to which they felt prepared for lectures. These results suggest that the new teaching method, although more engaging, did not increase students’ motivation to study the material independently. We note, however, that the time spent by UK university students on independent study has generally declined in recent years (Neves & Hillman, 2017) and that the validity of self-reported time use in retrospective student surveys is open to question (Porter, 2011). We also note that the number of minutes of online lectures watched by each student was poorly correlated with their final exam result (data not shown). Although one might conclude that the interactive problem-solving sessions were more pedagogically effective than the online component of the redesigned course, when asked to “indicate the aspects of this course that most helped you increase your knowledge and skills” in the free response part of the course evaluation survey, students mentioned the online components roughly twice as often as the problem-solving sessions. The perceived value of the online lectures and quizzes was therefore considerable.

There is no evidence that the change in teaching method boosted the exam performance of the students in BS1030 relative to the same students’ performance in BS1090. A general linear model of student performance (data not shown) suggested that an improvement in students’ incoming qualifications from 2014-15 to 2015-16 was probably responsible for the decreased failure rate in both courses over this period. A change from negative marking to number-right marking of MCQs seemed to be responsible for the increase in marks from 2015-16 to 2016-17 in both courses, but particularly so in BS1090, where no widespread change in teaching method was adopted.

The most rigorous meta-analysis of active learning effects in undergraduate science classes (Freeman *et al.*, 2014) suggested that the use of such methods might be expected to boost students’ examination performance by 0.47 standard deviations (SD). Using the lowest standard deviation measured in this work for BS1030 (SD = 13.3% in 2016-17) this equates to a predicted increase in the mean exam grade of just over 6% from before the intervention (2013-14 and 2014-15) to after the intervention (2015-16 and 2016-17). In fact the mean grade hardly changed, inching up from 50% in 2014-15 to 51% in 2016-17. (In contrast, the mean grade in BS1090 increased from 44% to 56% in the same period). The significant reduction in failure rate over the same period, owing to the decreased standard deviation visible in Figure 4, was seen to almost exactly the same extent in BS1090. Although the control in this experiment is by no means perfect – BS1090 is taught by various academics, including the author, and changes in both personnel and teaching styles have likely improved the teaching on this course over the last few years – we can say with some confidence that the semi-flipped, interactive teaching method adopted in BS1030 has not produced a marked improvement of exam results. This is in line with some earlier studies of problem-based learning (Prince, 2004).

There are several possible explanations for the failure of this intervention to deliver an improvement in average exam grades. One is the great variety of published active learning interventions, which extends even to the vague definition of active learning itself (Prince, 2004). Another is the inherent difficulty in performing properly controlled educational experiments, and the impossibility of precise experimental replication (Campbell & Stanley, 1966). A third possibility is that, despite efforts to mitigate the file drawer problem (Freeman *et al.*, 2014), publishing bias has tended to overestimate the effect of certain kinds of active learning in the exam performance of university science students.

Evidence for this possibility, at least in 15-year-olds, comes from the most recent Programme for International Student Assessment (PISA), a triennial survey of knowledge and understanding administered to half a million schoolchildren in the Organisation for Economic Cooperation and Development (OECD). PISA 2015, which focused particularly on science education, asked pupils taking the test how often their teachers explained ideas to the class, led discussions, and gave demonstrations (OECD, 2016). The authors then compared children’s test scores with the frequency of their exposure to these teaching methods, accounting for the socio-economic status of both children and schools. The most effective teaching strategy, which was correlated with improved test performance, was that of the teacher regularly explaining ideas to the class: a passive teaching technique, analogous to a traditional university lecture. In contrast, children who reported frequent whole-class discussions scored slightly lower than average on the test. Frequent use of other active, enquiry-based teaching practices, such as asking children to design and carry out their own experiments, was found to be associated with even worse performance (OECD, 2016). The striking contrast between these results and those reported at university level (Freeman *et al.*, 2014) deserves investigation.

**FUNDING**

This work was supported by Royal Holloway, University of London *via* a College Teaching Initiative Grant and a College Teaching Fellowship to the author.

**ACKNOWLEDGEMENTS**

We thank James March for his work as a BS1030 teaching assistant in 2016-17, and Leon Lloyd and Ben Hutchens of Pearson for their technical support of Learning Catalytics.

**REFERENCES**

Campbell DT & Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin, 1966.

Crowe J & Bradshaw T. *Chemistry for the Biosciences, 3rd ed.* Oxford: Oxford University Press, 2014.

Eichler J & Peeples J. Flipped classroom modules for large enrollment general chemistry courses: a low barrier approach to increase active learning and improve student grades. *Chem Educ Res Pract* 2016;**17**:197-208.

Felder RM & Brent R. *Teaching and Learning STEM: A Practical Guide*. San Francisco: Wiley, 2016.

Freeman S, Eddy SL, McDonough M*, et al.* Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci U S A* 2014;**111**:8410-5.

Haak DC, HilleRisLambers J, Pitre E*, et al.* Increased structure and active learning reduce the achievement gap in introductory biology. *Science* 2011;**332**:1213-6.

Julie S, Brian L & Eric M. Catalyzing learner engagement using cutting-edge classroom response systems in higher education. In *Increasing Student Engagement and Retention Using Classroom Technologies: Classroom Response Systems and Mediated Discourse Technologies,* Vol. 6 Part E (Wankel C & Blessinger P, eds.), pp. 233-61. Bingley: Emerald, 2013.

McClean S & Crowe W. Making room for interactivity: using the cloud-based audience response system Nearpod to enhance engagement in lectures. *FEMS Microbiol Lett* 2017;**364**:1-7.

Michael J. Where’s the evidence that active learning works? *Adv Physiol Educ* 2006;**30**:159-67.

Mooring SR, Mitchell CE & Burrows NL. Evaluation of a flipped, large-enrollment organic chemistry course on student attitude and achievement. *J Chem Educ* 2016;**93**:1972-83.

Neves J & Hillman N (2017) Student Academic Experience Survey. Higher Education Academy, York.

OECD. *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: OECD, 2016.

Pearson. *Learning Catalytics*, 2017. [www.learningcatalytics.com](http://www.learningcatalytics.com) accessed 21st July 2017.

Phillips JL. Do students think as we do? Progress with Piaget. *Improv Coll Univ Teach* 1982;**30**:154-8.

Porter SR. Do college student surveys have any validity? *Rev High Educ* 2011;**35**:45-76.

Preszler RW. Replacing lecture with peer-led workshops improves student learning. *CBE-Life Sci Educ* 2009;**8**:182-92.

Prince M. Does active learning work ? A review of the research. *J Eng Educ* 2004;**93**:223-31.

Rinaldi VD, Lorr NA & Williams K. Evaluating a technology supported interactive response system during the laboratory section of a histology course. *Anat Sci Educ* 2016;**10**:328-38.

Stockwell BR, Stockwell MS & Jiang E. Group problem solving in class improves undergraduate learning. *ACS Cent Sci* 2017;**3**:614-20.

Taber KS. Constructivism in education: Interpretations and criticisms from science education. In *Handbook of Research on Applied Learning Theory and Design in Modern Education,* Vol. 1 (Railean E, Walker G, Elçi A & Jackson L, eds.), pp. 116-44. Hershey: IGI Global, 2016.

**Scheme 1**

****

**Scheme 1.** BS1030 teaching design before and after the active learning intervention. Top panel: before the active learning intervention, in 2013-14 and 2014-15. Bottom panel: after the intervention, in 2015-16 and 2016-17.

Figure 1

****

**Figure 1** Screenshot of the summarized and anonymized class answers to an example “region” type Learning Catalytics question used in the course. The question was “Click on the amine group in this amino acid, arginine, that will react to form a peptide bond in the ribosome.” Green circles indicate correct answers, red circles the incorrect answers. The text below the diagram records the students who signalled their understanding, or lack of it, during the class discussion.

FIGURE 2

****

**Figure 2** Mean course grade achieved by BS1030 students in 2015-16 and 2016-17 (*N* = 228) who participated in different numbers of Learning Catalytics problem-solving sessions. Error bars represent standard errors of the mean.

FIGURE 3

****

**Figure 3** Likert scale responses to two BS1030 course evaluation survey statements: “I received helpful feedback on my progress during the course” (left panel) and “The teaching style was engaging” (right panel). A response of 1 indicated strong agreement, and 5 strong disagreement. No responses of 4 or 5 were obtained. The average number of responses to each question was *N* = 65.

FIGURE 4

****

**Figure 4** Boxplot of final exam marks in BS1030 (dark grey) and BS1090 (light grey) over four academic years: two years before the active learning intervention (left of the dashed line) and two years after the intervention (right of the dashed line). Each box represents an interquartile range (IQ), and the horizontal line in each box is the median mark. The “whiskers” extend 1.5 IQ above the 3rd quartile and below the 1st quartile, and the open circles represent outliers beyond this range.

GRAPHICAL ABSTRACT

****