The inverted classroom in a large enrolment introductory physics course: a case study.

Prof Simon Bates & Dr Ross Galloway
School of Physics and Astronomy
University of Edinburgh
Edinburgh EH93JZ
S.P.Bates@ed.ac.uk ross.galloway@ed.ac.uk

Abstract
We present a practice-based case study of curriculum redesign in a large-enrolment introductory physics course taught at the University of Edinburgh. The course has been inverted, or "flipped", in the sense that content and material is delivered to students for self-study in advance of lectures, via a combination of home-grown electronic course materials, textbook reading and external web resources. Subsequent lectures focus on problems students are still having after self-study of the material, which have been self-reported by them as part of a weekly reading quiz assignment. Lectures are transformed from sessions for transmission or initial presentation of information, to guided discussion sessions, with a particular focus on peer instruction techniques and discussion, facilitated by extensive use of clicker questions. We present details of student engagement with pre-class reading and quiz tasks, comment on student perceptions of this different instructional format, and present data that shows evidence for high quality learning on the course.

Keywords
Physics, curriculum reform, clickers, peer instruction

1. Introduction

In STEM subjects, and indeed many others, lectures are still a major component of most undergraduate courses. They are efficient but not particularly effective vehicles for promoting deep student learning. There has been much written about techniques to make them more effective teaching and learning experiences (Bligh, 2000), with probably the most well-documented engagement technique introduced in the past decade being the widespread introduction of electronic voting systems into (principally) large class lectures (Bruff, 2009). Notwithstanding these enhancements, the progress away from lectures as a mechanism for transmitting information is far from complete.

In this paper, we present details of a major curriculum reform to our first year introductory physics courses at the University of Edinburgh, with a particular focus on redefining the role of lectures as whole class meetings. The changes introduced in the present academic year are based around a pair of well-documented teaching methodologies: Just-in-Time-Teaching (Novak et al, 1999) and Peer Instruction (PI) (Mazur, 1997), both of which have disciplinary roots within physics. The former involves students completing outside-class reading and assignments, the results of which are used by the instructor to influence and inform the direction of subsequent teaching sessions such as lectures. The latter is an in-class methodology developed to promote student discussion and learning in lectures, based on discussions around conceptual questions posed by the instructor.

The combination of the two techniques has recently been referred to as inverting, or "flipping" the classroom structure: moving content coverage outside the classroom, in order to spend precious in-class time on more demanding tasks. In terms of the cognitive domain of thinking skills, such as those defined by Bloom's (revised) taxonomy (Krathwohl, 2002), the notion of "flipping" is synonymous with using lecture time to visit the upper regions of the taxonomy. Bloom's taxonomy is often represented in pyramidal form, with activities such as analysis and synthesis at the highest levels, to be compared with a model of lectures as vehicles for information transmission, firmly rooted to the lower regions of the pyramid (if there is any meaningful cognitive activity at all in such sessions).

This paper describes our experiences with the planning and delivery of this course. We begin with some necessary contextual information about the class and course cohort, and describe the weekly rhythm of teaching activities. We then present details of student engagement with the process, their feedback and views on the course structure and delivery and also present evidence for significant learning during the course. We conclude with a practical and pragmatic look at logistics, including workload implications for staff involved.

2. Course details and "inverted" methodology

"Physics 1A: Foundations" is a first year, first semester course in classical mechanics and dynamics at the University of Edinburgh, compulsory for all students on Physics degree programmes (which in Scotland are generally of 4 years duration) but also taken as an elective by students on a wide variety of others. These two sub-populations are approximately equally represented in a total course cohort of around 200. The course has a long history of introducing and evaluating a variety of innovative teaching methods: details of how online resources are used to complement face-to-face teaching have been presented elsewhere (Bates et al, 2005), as has the role of studio-based workshops for development of collaborative problem solving skills in novel teaching spaces (Bates, 2005). Most recently, we have introduced student-created questions as a component of the summative course assessment (Bates et al, 2011).

Whole class lectures (1 hour duration) take place three times a week. These are supplemented by 3 hour weekly workshop sessions, in which students develop a range of skills, both discipline specific and also more general (Bates, 2005). The course does not have a laboratory component, but the second semester companion course, which is taken by virtually all 1A students, does have a full laboratory programme. This year, we set weekly reading targets for students, in which they were expected to cover material that would be the subject of the following week's lectures. As these were first year students, we provided guidance as to what exactly we expected of them during this self-study (i.e. much more than simply glancing over the material and thinking "that looks familiar: I can do that"). Accompanying the weekly reading was a weekly reading quiz, delivered online through the course VLE, which needed to be completed by 8am on a Monday morning. The quiz comprised 6 questions: the first five were straightforward items testing understanding of content and the student score on these comprised a small component of in-course assessment (1% per reading quiz, 10% over the whole semester). The final question was an unscored free text response to the statement: *"After completing this week's reading, what I still don't understand is….".* These responses were collated at the start of the following week, categorized and used as the basis for what material would be focused on in lectures.

Lecture sessions were built around a series of clicker questions, many of which were used as the basis for peer instruction. The typical running order of a PI session is as follows: the instructor poses a conceptual question; student votes individually using clicker handsets; students are not shown the initial response graph immediately and if the proportion of correct responses is suitable (usually between >30% and <70%), the class is invited to find someone close to them who voted differently to them and engage in a discussion. In our lectures, we have specifically encouraged students to move if they find themselves in an enclave of similar thinkers. After a few minutes, a revote is taken, followed by further instructor led discussion if necessary. All lectures contained one or more episodes like this, but other activities were also undertaken, such as working through examples or concepts. The emphasis was very much on participatory discussion *with* the class, rather than instructor presentation *to* the class.

3. Results

In this section, we present our results from the delivery of the course, separated into considerations of student engagement, student feedback and student learning. Data tables and graphs can be accessed as supplementary information online (Bates and Galloway, 2012).

3.1 Evidence for student engagement

Students did engage with the weekly reading quizzes that were set for them each week. Participation rates ranged from a minimum of 83% (first and last weeks of the course) to a maximum of 95% (several weeks). The average participation rate (± 1 standard deviation) over all ten quizzes was 91%(±5), N=199. Furthermore, they engaged with the quizzes *meaningfully,* with the cohort average score per quiz ranging from 70% (minimum) to 90%

(maximum), with a mean of 79%(±6). As might have been expected, the vast majority of student submissions to the reading quiz were made in the 48 hours immediately prior to the submission deadline. The free response question, asking students to identify what they still did not understand after completing the reading and quiz, was answered by approximately 70% of those completing the quiz: virtually all submissions were relevant to the material in the course. Collated student submissions to each weekly occurrence of this question ran to several pages of commentary, and were a valuable source of feed-forward to aid lecturing staff to really get a sense of the topics causing students genuine difficulty. Selected but typical comments from students included:

"Newton's laws seem a lot more complicated than the simplistic versions i was taught in school."

"For me electrostatic forces are the most difficult in this weeks reading. Especially doing vector calculations with charges in an electric field."

"I found the idea of the centre of mass to be interesting, as while it has been implicitly used in my previous physics courses, it has never been explicitly defined."

As well as this quiz item providing valuable feedback to lecturing staff to determine the topics for lecture coverage, encouraging this metacognitive reflection by students on what topics they were having difficulties with is also expected to be beneficial to their own learning.

An often-heard comment relating to provision of material to students (usually lecture notes) in advance of class sessions is "If you give them the lecture notes, they might not / won't turn up". We gave students not just lecture notes, but in effect the entire course content in advance of class sessions: it might reasonably be asked did we not have empty lecture theatres by week 5? In fact, we did not see any evidence of a significant decline in lecture attendance[1], which we were able to "measure" by observing a relatively constant number of total clicker votes per question (across 140 individual clicker question episodes) as a function of a time period spanning 11 weeks of the course. There was a slight decline towards the final week of teaching in the semester, perhaps partly explained by the effects of a long teaching semester taking its toll and the looming shadow of degree examinations 2 weeks after the course concludes. This teaching methodology, therefore, provides evidence against the 'no notes in advance' argument as a technique to maintain student attendance and engagement.

[1] That is not to say our lectures achieved 100% attendance. In actual fact the average attendance was around 70-80% (there are a whole host of possible reasons for this which we do not go in to here). What we did observe is a relatively constant attendance over time.

3.2 Student feedback

With a raft of changes introduced into the course delivery this year, we were particularly attentive to collecting student feedback on the course and its delivery, both during and after the course teaching. Clickers were used both for subject-based conceptual tests and to gather class feedback in lectures. All of our end of module questionnaires contain the same "bottom line" question and we additionally asked a question specifically about the style of course delivery. Responses are summarised in Table 1. Free text responses collected from the same questionnaire indicated a high level of student satisfaction with the course design and delivery, even if it took them some time to adjust to the methods used and expectations. The following quotes illustrate this:

"*The style took some time to grow on me, but I now prefer it*"

"*Have decided I am really not a fan of the traditional "take notes from a slideshow" lecture – not much thinking ends up being done and it makes it far too tempting to skip the lecture altogether and just pick the notes up online."*

Question text	Poor	Weak	Average	Good	Excellent
"Overall I feel this course was…"	0	1	3	45	51
	Strongly prefer this approach	Slightly prefer this approach	Don't mind either way	Slightly prefer traditional approach	Strongly prefer traditional approach
"This course was delivered in a different style to more traditional lecture courses. Please indicate your preference."	54	28	10	6	2

Table 1: Percentage of responses to selected questionnaire items. For both items, N=97 (49%)

3.3 Evidence for learning

For the last 6 years, we have been measuring student learning gains on the course, using the Force Concept Inventory (FCI) (Hestenes et al, 1992). This diagnostic instrument has been deployed to hundreds of thousands of students studying introductory Physics courses all over the world, and is in many ways the "gold standard" conceptual inventory in the physical sciences, providing a rigorous examination of the extent of students' conceptual understanding of the Newtonian concept of force. The standard deployment protocol (Bates and Galloway, 2010) is to ask students to take the test prior to any teaching ("pre-test"), and then at the end of the teaching relating to that section of the course ("post-test"), and to calculate the average gain (improvement) for the cohort. As students often start from very different pre-scores, it is normal to calculate a *normalised gain*, i.e. improvement as a fraction of total possible improvement. An analysis of FCI normalised gains (Hake, 1998) has been used to evaluate the effectiveness of different teaching methodologies in improving conceptual understanding. Normalised gains of 0.3 (meaning on average, a typical student improves their mark by one-third of what they could maximally improve) are generally indicative of effective interactive engagement courses, with many courses achieving gains of 0.5 or even higher. Based on our FCI results for this cohort, the class recorded a normalised gain of 0.54, which is towards the higher end of what we have seen looking back over a period of years. Most notably, the modal score on the post-test was 100%, and the cohort average was 85.4%, above what is classed as the threshold for "Newtonian mastery" of the force concept. The histograms for the pre- and post- FCI tests are shown in Figure 1.

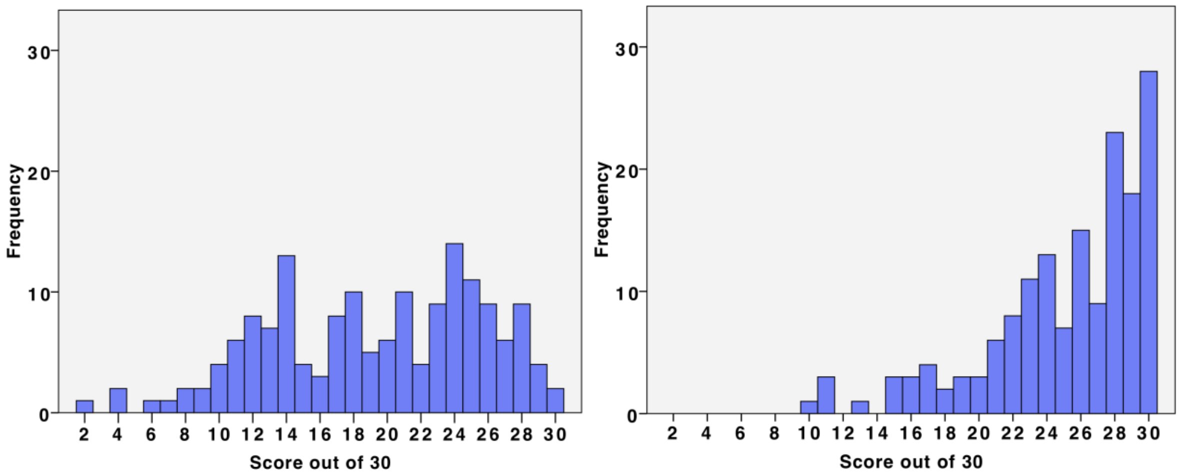

Figure 1: Pre (left) and post (right) instruction histograms of FCI score. Student data was matched (ie only students completing both pre and post tests were included, *N*=161).

The large gains measured by the FCI scores show significant learning has taken place through all formal and informal teaching episodes on the course. Focussing in on learning in lecture sessions, we can apply a similar analysis to calculate a normalised gain for the PI question pairs that form the basis of the discussions in lectures. As stated previously, not all clicker questions are suitable for productive PI discussion, depending on the proportions of correct responses in the initial individual voting on the question. Clearly, if either proportionately too few or too many students initially choose the correct answer, the probability of finding someone with a different answer choice to promote effective discussion is low.

During the course, we completed 41 PI sessions during lectures. These sessions account for approximately 50% of the clicker questions posed in lectures. In this case, the calculation of the normalised gain is the fraction of students improving (i.e. getting the particular PI question correct) as a fraction of those who initially got it incorrect. Of the 41 PI question pairs, 40 of them have positive normalised gains, with an average gain of 0.45 (±0.22). In some cases, very high gains (>0.7) were observed, occasionally as high as 0.9. Figure 2 shows the normalised gain profile for all 41 PI sessions.

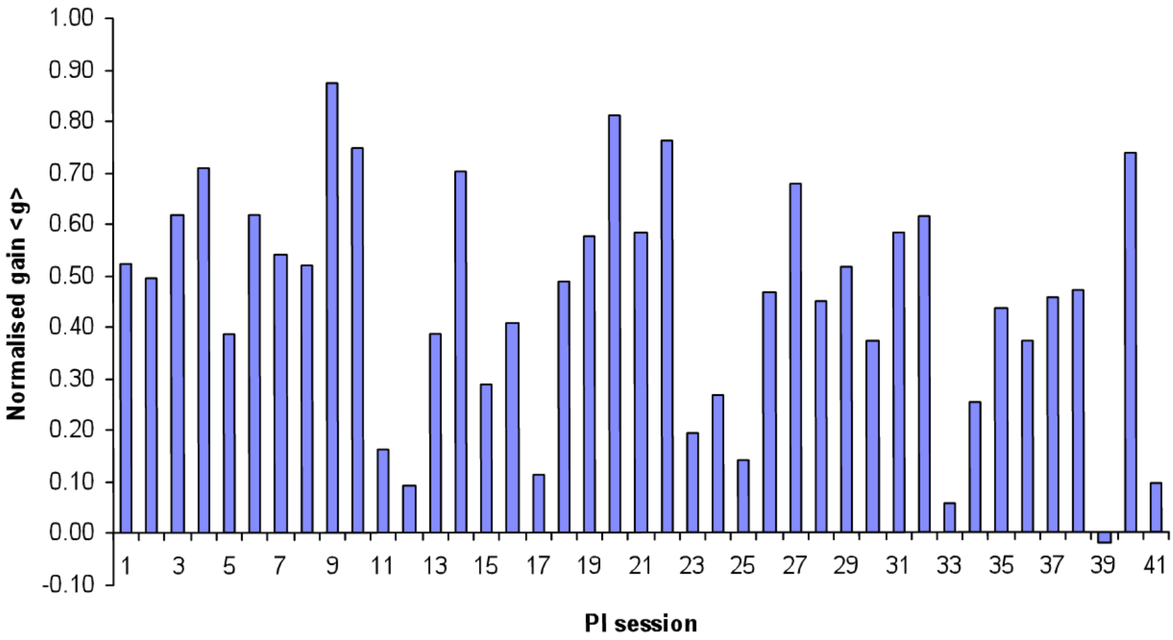

Figure 2: Normalised gain on PI discussion question pairs.

4. Discussion

By all measures we have used to evaluate it, the course presentation this year has been a resounding success. This has been confirmed by the end of course examination results. This year, the examination format was changed to an 'open note' examination (students could take in their course handbook, issued to them at the start of the course, and any other notes they have created themselves). The examination focused heavily on problem solving rather than bookwork. The pass rate this year was 89%, with an average mark at the upper end of recent paper averages.

An obvious concern, however, is that of staff workload: did this not necessitate a huge up-front investment of staff time? There are indeed additional up-front costs, for example we had to prepare 10 weekly quizzes. There is also the requirement to have access to a large quantity of good quality clicker questions to form the basis of lecture discussions, or else to incur the cost of creating these. This is no longer the huge overhead it once was. Many introductory course texts come with a database of such questions, and online repositories and user groups around the world mean that it is probably not necessary to start from scratch in any STEM subjects anymore. In an effort to manage workload, we delegated the task of collating weekly free-response comments from students doing the reading quiz to one of our postgraduate teaching assistants, saving us an hour a week as course lecturers.

Perhaps more significant than the additional workload is the mental shift that is required to accept and embrace an unstructured, contingent lecture experience in which the lecturer is no longer in complete control of. Carefully planned lecture timings, where one might like to know that one will get to the bottom of page 5 by the halfway point of a particular lecture, need to be abandoned. The payoff for this letting go of complete control is the potential for an inclusive and participatory classroom atmosphere: we managed to create a genuinely open discussion involving 200 people. We spent as much time halfway up the raked steps of the lecture theatre as behind the bench at the front. We created an environment, largely through clear expectation setting from the very first lecture, where people would volunteer comments and suggestions, not just from "the usual suspects", and without being crippled by the fear of getting something wrong. It is an exhilarating feeling to be freed from the tyranny of content coverage to be able to have the time and space to focus on what really matters: whether or not students actually understand the material. Furthermore, this understanding did not come with a price of "covering" less material: we are convinced that, largely through the students' efforts outside class, we covered as much content but uncovered a great deal more understanding.

5. References

Bates S.P. and Galloway R.K., (2012) *The Inverted Classroom: what it is, why we need it and what it might look like*, Prezi.com. http://bit.ly/invertedclassroom [accessed 3rd February 2012].

Bates S.P., Galloway R.K., McBride K.L., (2011) *Student generated content: using PeerWise to enhance engagement and outcomes in introductory physics courses*. Proceedings of the 2011 Physics Education Research Conference, To appear.

Bates S.P., Galloway R.K., (2010) *Diagnostic tests in the Physical Sciences: a brief review.* New Directions: the Journal of the Higher Education Academy Physical Sciences Centre 6, 10-20.

Bates S.P., Bruce A.D., McKain D. (2005) *Integrating e-learning and on-campus teaching I: an overview*. Research Proceedings of the 12th Association of Learning Technology Conference ALT-C 2005 'Exploring the frontiers of elearning' Manchester, p130-139.

Bates S.P., (2005) *Reshaping large-class undergraduate science courses: the weekly workshop.* CAL-laborate, 14 1-6.

Bligh, D. (2000), *What's the use of lectures?* 6th Edition, Wiley and Sons, UK

Bruff, D. (2009), *Teaching with Classroom Response Systems,* 1st Edition, Jossey-Bass, San Fransisco.

Hake R.R., (1998) *Interactive Engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses,* Am. J. Phys., 66(1), 64-74.

Hestenes D., Wells M., Swackhamer G., (1992) *Force Concept Inventory,* The Physics Teacher, 30, 141-158.

Krathwohl, D.R., (2002) A revision of Bloom's taxonomy: An overview, *Theory into Practice,* 41, 212–218.

Mazur, E. (1997). *Peer Instruction: A User's Manual*, Prentice Hall, Englewood Cliffs.

Novak G.M., Patterson E.T., Gavrin A.D., Christian W., (1999), *Just in Time Teaching,* Prentice-Hall, New Jersey.