Evidence for the Flipped Classroom in STEM

Mike Dodds
mike.dodds@york.ac.uk

July 2015

(I wrote this survey paper for York's PGCAP teaching course. I'm not an education expert, but if you're considering flipping a STEM classroom, it might be useful. – Mike)

Question: To what extent is there evidence that so-called 'flipped' learning techniques improve student learning outcomes, and how can flipping be applied in a Computer Science curriculum?

1 Survey Aims

The *flipped classroom* (or *flipping*) [4] is a new educational technique which seeks to invert the traditional model of in-classroom lectures and out-of-class homework. Instead, it advocates in-class interactive group learning, and out-of-class instruction via videos or podcasts. Intuitively, this approach moves classroom time away from mere dissemination of information, and allows teachers to focus on solving student problems and enabling learning. This connects with the increasing evidence that problem-based learning is an effective teaching technique (see e.g. the meta-analysis in [9]).

The intuition for flipping seems compelling, reflected in attention in the non-academic media (e.g. [11, 18]). However, flipping is not trivial to apply. Applying it means restructuring the entire course to focus on group learning, while also recording supporting instructional material. Given the up-front costs, it is important to know whether flipping is truly effective before applying it in practice.

In this study, I investigate the evidence that flipping is effective in improving student learning outcomes in STEM subjects, and I also examine some of the pitfalls and opportunities flipping presents. My aim is to understand better how these ideas can be applied to teaching in my home department, Computer Science.

Survey aims. My concrete aims for the survey were as follows:

- To review the education literature on the flipped classroom when applied to STEM subjects. As far as possible, I surveyed the literature in a rigorous manner, focusing on empirical research which directly measures student learning outcomes.
- To derive broad conclusions regarding the state of research on flipping. In particular, what aspects of flipping have been studied, and what practices are supported by evidence?

- To sketch a flipped redesign of a module in my home department, Computer Science, taking evidence from the education literature into account. My aim was that this hypothetical redesign would both motivate this survey, and help me understand the specific context in my department.

2 Research Method

Given my expertise and the amount of time available, this report is not a systematic survey of the literature. However, in order to make my review as rigorous as possible, I took several steps to fix a research method ahead of time: I chose a clear definition of flipping; based on this, I fixed a criterion for inclusion in my review; finally, I defined a method for locating research papers.

Defining 'flipping'. I will follow Bishop and Verleger's prior survey paper [5]:

> *"We define the flipped classroom as [consisting] of two parts: interactive group learning activities inside the classroom, and direct computer-based individual instruction outside the classroom."*

Flipping is perhaps less radical than it initially appears. Group-based problem classes as are widespread in STEM teaching, including York Computer Science. As noted in [5], this kind of group activity derives from long-established student-centered learning theories such as [21]. Furthermore, modules where preparatory reading is expected are also common: for example, reading groups or seminars in Humanities subjects.

In the academic education literature, I found that 'flipping' is now frequently used in a broader sense to mean any teaching strategy where core material is disseminated independently (e.g. through textbook reading) and class time is primarily spent on group problems or interaction with the teacher. However, my definition requires that external teaching is *computer-based*, which rules out this broader sense of 'flipping'. This allowed me to focus specifically on techniques that have been enabled by new technology. It also helped keep the scope of my work under control, and allowed some measure of comparability between studies.

Inclusion criteria. I fixed the following criteria for consideration in my reading:

1. *The technique studied should meet the above definition of flipping.*
2. *The topic taught should be a STEM subject.* Restricting my reading to computer science alone would have left very few papers. My intuition was that STEM subjects share similar teaching approaches and problems (e.g. when compared to Humanities subjects).
3. *The research should consist of an empirical trial, not just an experience report.* The large amount of attention it is receiving shows that flipping is intuitively compelling, but I was interested in understanding the claims about practices that are supported by evidence.
4. *The research should compare student learning to a control group, not just survey student perceptions.* I made this decision for two reasons. First, student learning,

rather than enjoyment, is the key aim of higher education. Second, the literature appears to be in agreement that students have good perceptions of flipped classrooms, and I wanted to study a question with a potentially less settled result.

5. *The study had to be conducted in a real-life classroom or programmatic setting.*

Several of these points were inspired by [9], which describes a meta-analysis of research into problem-based learning. (Note that unlike [9] my survey is not a statistically rigorous meta-analysis).

Literature search method. My starting point for locating literature was the Education Resources Information Center (ERIC) search engine [3]. I chose ERIC because it was used in the meta-analysis of problem-based learning [9], and because it is a long-standing source of educational resources backed by the US government. I searched ERIC for the string "flipped OR flipping", and filtered for peer-reviewed papers published in the last five years (since 2011). This yielded 73 papers. I then filtered by hand according to the criteria given above, yielding 10 papers.

For papers earlier than 2011, I used ERIC, combined with references gleaned from [5], a survey paper on flipping from 2012. This yielded a further 2 papers. I also looked in the identified papers for further relevant references ('snowballing'). This only identified one further paper, [12], for a total of 13 papers. This fact, together with a brief search of other education sources, leads me to believe I achieved a good coverage of the current education literature.

3 Results From Identified Literature

The papers which fit my inclusion criteria were [6, 8, 7, 10, 12, 13, 15, 16, 17, 19, 20, 22, 23]. See Figure 1 for titles – full references in Bibliography. In 2012, the survey by Bishop and Verleger [5] identified only two studies of flipping which satisfied my criteria – [8, 17]. The first observation to be made is therefore that research on flipping has become much more active over the last half decade, even within the restricted parameters I have set.

Performance and perception effects. All of the studies but two report significant improvements in student performance with respect to the chosen control group. (Note that the studies vary as to whether they tested these improvements for *statistical* significance – see below). The dissenting studies are Clark [6] and Jensen [12], both of which report a slight, statistically non-significant improvement in performance. It therefore seems reasonable to conclude that *flipped classrooms are an effective teaching technique in comparison to lecture-based learning.*

All but one of the studies surveyed student perceptions of flipping (the exception was [23]). Fewer of the studies compared these results with a control group – this is because many of the studies compared with previous years, for which presumably no data on perceptions existed. However, all the studies which measured it report positive perceptions from students. It seems reasonable to further conclude that *students respond positively to flipped techniques.*

Taken together these two results seem a strong argument in favour of adopting flipped techniques in STEM teaching.

Ref.	Title	Subject
[6]	*The effects of the flipped model of instruction on student engagement and performance in the secondary mathematics classroom*, K R. Clark.	Maths
[7]	*Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course*, R. S. Davies, D. L. Dean, and N. Ball.	Computer Science
[8]	*Evaluating a Web Lecture Intervention in a Human Computer Interaction Course*, J. A. Day and J. D. Foley.	Computer science
[10]	*The flipped classroom for teaching organic chemistry in small classes: is it effective?*, J. M. Fautch.	Chemistry
[12]	*Improvements from a flipped classroom may simply be the fruits of active learning*, J. L. Jensen, T. A. Kummer, P. D. d. M. Godoy.	Biology
[13]	*Flipped Physics*, M. Kettle.	Physics
[15]	*Student learning and perceptions in a flipped linear algebra course*, B. Love, A. Hodge, N. Grandgenett, A. W. Swift.	Maths – algebra
[16]	*Flipping calculus*, J. McGivney-Burelle and F. Xue.	Maths – calculus
[17]	*Learn before lecture: A strategy that improves learning outcomes in a large introductory biology class*, M. Moravec, A. Williams, N. Aguilar-Roca, D. K. O'Dowd.	Biology
[19]	*Flipping the Classroom and student performance in advanced statistics: evidence in a quasi-experiment*, M. Touchton.	Maths – statistics
[20]	*Flipped classroom model improves graduate student performance in cardiovascular, respiritory, and renal physiology*, J. D. Tune, M. Sturek and D. P. Basile.	Physiology
[22]	*The flipped class: a method to address the challenges of an undergraduate statistics course*, S. G. Wilson.	Maths – statistics
[23]	*Flipped statistics class results: better performance than lecture over one year later*, J. R. Winquist and K. A. Carlson.	Maths – statistics

Figure 1: Identified papers from literature survey – see Bibliography for full citations.

Criticisms. Several criticisms can be raised of these conclusions.

To begin with, the design of studies varied considerably, in terms of the topic studied, length, precise classroom practice, control groups, and others. It would therefore be a mistake to apply any 'flipped' technique blindly without considering exactly what approaches have and have not been studied. To help ensure that the techniques adopted are supported by evidence, teachers should look at the specific practices that have been studied. In §5 below when describing a flipped course redesign, I recommend using [19] as a direct model.

Jensen et al. [12] raise another criticism: gains from flipping may not be the result of the flip *per se*. They compare flipping with problem-based learning, and show that both exam outcomes and deep learning were similar with both approaches. This study raises the possibility that flipping is just one route to the real source of educational improvement: active learning. The same gains may be available through other routes that do not require so radically altering educational practice. However, flipping may nonetheless be an effective way to make time for active learning in the classroom.

Another concern raised by Clark et al. [6] is the possibility that students are responding to the *novelty effect*. This is described in [6] as *"the tendency for performance to initially improve when technology is instituted"*, for example with interactive whiteboards [14]. Prior exposure of students to traditional lecturing could make flipping appear more novel and exciting, amplifying its effect. This is related to the *Hawthorne effect*, where results improve during a study simply as a result of being observed. To achieve a fairer comparison, flipping would need to either be studied over a longer period, or tested with students who had not previously experienced traditional lecture-based learning.

However, I note that Davies et al [7] somewhat contradicts this concern: in this study, flipping and traditional lecturing were compared with a third strategy, *simulation-based teaching*. Broadly, flipping performed better than lecturing, while simulation performed worse despite its novelty. In this case, the novelty effect alone cannot be responsible for flipping's success.

A third concern is that none of the studies rigorously examine the workload cost of flipping. Firstly, this makes them less useful when deciding whether to implement flipping: it is difficult to know whether costs would be reasonable relative to gains. However, this also makes it harder to know whether flipping is truly effective. It would be unsurprising to discover that increased teacher effort translated into improved student learning. If flipping is no more effective than just spending more teaching resources on students, there is little argument for applying it.

In fact, the papers I have identified generally *do* require increased effort relative to traditional lecturing. This is because teachers both prepare pre-recorded material and lead problem classes (the exception is [22], which used material from Khan Academy). Several of the studies comment that recorded lectures require more preparation than delivering them in the traditional manner. The (largely unstated) assumption is that material prepared for flipped courses can be reused in future iterations of the course, amortising this cost. However, this assumption is not tested, and as course material is often revised between years, there is some reason for caution.

Differences between studies. Most importantly, the learning activities themselves varied. In most of the studies, video lectures were prepared by the teacher. However, [22] used Khan Academy videos and the textbook as the source of out-of-class learning. In most of the studies, the primary in-class activity consisted of group-based problems, and discussion with the teacher. However, there were also variations in this: for example [20] devoted class time to a quiz and discussion. Quizzes at the start of class were a common strategy aimed at ensuring students completed preparatory work.

The topic studied also varied considerably, with a noticeable skew towards mathematics. As Figure 1 shows, six of the 13 studies focus on maths topics, with four focusing specifically on statistics. It is unclear whether how this skew should be interpreted – it may be that maths courses are already taught through problem classes, making them particularly suitable to flipping. However, the positive results for flipping should be treated cautiously when applied outside this domain.

The control groups used in the evaluation varied in their nature. Six of the studies compared results with a previous year's unflipped course [6, 10, 13, 17, 23]. One problem with this is that redesigned courses might cover substantially different material. This approach also precludes comparing student attitudes, as these were not surveyed in previous years. Finally, it presents the problem that student groups could differ across years as a result of factors not accounted for in the study. The other studies ran concurrent flipped and non-flipped modules [8, 7, 12, 15, 16, 19, 20]. Informally, these studies also took more care that control group and study group were taught the same material.

The length of the flipped and non-flipped courses themselves varied. Most of the studies covered either a whole course or a substantial portion of a larger course. In most studies learning outcomes were examined immediately after the course. An interesting exception is Winquist et al [23]. In this study, learning outcomes were examined a year after the flipped portion of the course had finished. Students demonstrated improved learning outcomes with respect to their non-flipped contemporaries, and with respect to other non-flipped portions of the course. Comparing student groups both with others, and with their own performance on other portions of the course seems like a good approach.

Although I have grouped together the control groups under the term 'traditional lecturing', inevitably the nature of these activities varied considerably as well. An extreme example is Jensen [12], which deliberately compares flipping with a course based on class discussion and problem-based learning. This is very different from a stock lectures-and-practicals design. Similarly, a large number of practices fall under the umbrella of 'traditional lecturing' in my home department. For example, some lectures involve in-class problems and discussion, some are interspersed with problem classes, for some attendence is voluntary, and so on.

The studies also varied in the statistical sophistication of their analysis of student learning outcomes. Three of the studies simply reported changes in grades, and did not check their results for statistical significance [10, 13, 16]. The remainder did so, and several also statistically analysed change in student perceptions, where the data was available.

4 Lessons on Flipping

Lessons from the literature. Broadly, the following claims appear to be well-supported by empirical evidence:

- Flipping is an effective teaching technique which can improve student performance.
- Student feedback from flipped classes shows that students find it a useful learning approach.

The following lessons are not empirically tested, but are common comments in the papers I have studied:

- The initial workload cost of flipping is substantial, in designing the course, in preparing material, in mastering the technology needed, and in teaching assistant time.
- An important part of flipping is ensuring that students study material outside the classroom. Quizzes which contribute to the students' final grade are an effective way of ensuring this.
- Flipping leads students to come to problem classes better-prepared, meaning lecturers can cover more challenging material.
- While flipping is generally popular, a minority of students view it negatively. This may be because flipping requires more active rather than passive engagement with a course.
- All of the studies I found were broadly positive in their view of flipping, even those that did not find a statistical effect for it. Thus we can tentatively conclude that flipping is appealing for education professionals (at least those interested in novel teaching techniques).

Some of the individual papers identified particular lessons from flipping. Davies et al [7] argue that flipping supports 'scalability': the possibility of increasing the number of students on the course while maintaining its structure. Touchton [19] and Winquist et al. [23] emphasise the importance of immediate feedback or testing for student learning. Winquist et al also identifies the *generation effect* – students learn better when they are forced to generate material rather than passively read it.

Open questions. Several questions are left open by the current research on flipping.

In their book advocating flipping, Bergmann and Sams claim a number of reasons to adopt it [4, chapter 4]. Several are inherently true given the structure of flipped classrooms, e.g. "flipping increases student-teacher interaction". Many of Bergmann and Sams' claims are about student perceptions, e.g. "flipping speaks the language of today's students". The studies I have identified broadly seem to support the idea that students are happy with flipping as a teaching style.

Two other claims directly address student performance: "flipping helps students of all abilities to excel" and "flipping helps struggling students". The literature broadly seems to support the former claim: students do seem to learn more effectively in flipped classrooms. The latter claim is important but less studied. Fautch [10] identifies a reduced drop-out rate, and observes that: *"this, arguably, is one of the main goals of using the flipped classroom as a pedagogy – to help those students who might otherwise drop the course."* More research is needed to see how different kinds of student are affected by flipping.

More study is also needed to determine whether the benefits of flipping are just a result of the 'novelty effect'. It is possible these benefits would disappear if flipping is rolled out across a curriculum. In the longer term, this will become clear if flipping becomes a popular teaching style. Examining this in a small study would require testing flipping on students without experience of standard lecture-based techniques.

Finally, more empirical study is needed of the workload implications of flipping. Touchton [19] observes that: *"shifting from a traditional to flipped classroom carries considerable start-up costs"* but that *"class preparation time fell to almost zero by the third iteration of the course"*. Anecdotally, workload is the major concern for my colleagues in Computer

Science when flipped techniques are proposed. Studying this will require a long-term study that could capture the benefits of pre-recording, as well as its initial costs.

5 Prototyping Flipping in York Computer Science

The literature I have reviewed indicates that flipping is an effective and popular teaching approach, although several questions remain unresolved. While there is a considerable cost to redesigning modules, the literature does seem to justify prototyping flipping within York Computer Science.

In this section I will sketch a redesigned version of Mathematical Foundations of Computer Science (MFCS) [1], a module taken by approximately 140 first-year undergraduates. I do not expect that my MFCS design will be implemented directly; rather I hope it will feed into the departmental discussion on how flipped techniques can be integrated into the curriculum.

Current practice. Traditional lectures are the most common way of disseminating information in York CS, but the department also uses a number of other teaching approaches. Firstly, problem / practical classes are widespread through the curriculum. The common format is that these take place after the material has been covered in lectures, and are supervised by the lecturer and some number of teaching assistants. Other courses are run as reading groups, with students working in small groups and most learning taking place in group discussion, while still others are run as group or individual projects, with teaching staff available for feedback as required.

York CS is currently beginning its first experiment in flipping. In September 2015, the department will offer its first module based on flipping: Designing and Maintaining Software (DAMS) [2], led by Dr. Louis Rose. This is a third-year module with 30-40 participants, which will replace an existing module taught as a reading group. Fitting with other comments in the literature, Rose says that initial workload has been substantial, but he expects it to decline in future years once pre-recorded lectures can be reused. He also says practical classes will be much more challenging than is standard in the department: as discussed above in §4, this is justified because students should be better-prepared and will have more time to work on problems.

Motivation and design. MFCS is an interesting target for a flipped redesign for the following reasons:

- It is a large first-year module, and thus a substantially different case-study from DAMS.
- Current lectures are minimally interactive, while problem classes are used extensively.
- It is a module I know well, having helped teach it in 2014/15.
- It is based on a stable body of core knowledge which is unlikely to change.
- It is a mathematics course – as noted above, maths is the best-studied topic for flipping.

One potential objection is that MFCS is a key part of the curriculum: if a redesign creates problems, these will negatively affect many other modules. To mitigate this risk, I propose that initially only a portion of the module be flipped (this was suggested by Moravec et al. [17]). As a two-term module, MFCS is already split between two lecturers, so it would be reasonable to split off a portion for flipping.

In the design of the module itself, I use Touchton's design for a statistic class as a template [19]. By doing this, I hope to mitigate risk by following an established pattern. I choose Touchton's paper as it is well-described, targets a mathematical subject, and achieves positive results. A further benefit of replication is that this would allow comparison of our outcomes with his.

I propose that the flipped portion of the course will consist of roughly the final half, covering formal languages and automata. The course will be structured as follows:

- Material from the existing MFCS will be pre-recorded and broken into 15 minute segments. Reusing previous material will help mitigate the cost of flipping.
- Problem classes will be increased in number to compensate for the lack of lectures. They would therefore double to 16 per term, rather than eight. One difficulty is that this would create the need for more teaching assistants.
- Existing problem sets will be extended to cover more advanced material, with students encouraged to work on problems in small groups.
- Auxiliary mini-lectures will be given as needed in problem classes on topics where students are experiencing problems. (This is advocated by Touchton).

Touchton devotes the final section of his module to a student research project based on statistical case-studies. It is more difficult to see how to apply this to MFCS, as the material is quite abstract. Nonetheless, this would be a good idea if an appropriate collection of projects could be constructed. Touchton observes: *"For me, this assignment was the most important of the course because students applied statistical techniques acquired in class to a real-world problem"*. For MFCS, I suggest the flipped course initially omits this to simplify the redesign. However, the course should be structured so that a project component could replace some of the short problem sets in future years.

Another practice advocated by many of the studies I've examined (but not Touchton) is the use small summative tests at the start of each problem class. This helps ensure that students keep abreast of the course material. For MFCS, the large number of students in the class means that this could impose a substantial marking burden, so I choose not to include it in this design.

Evaluation. As the course material remains the same, it will be relatively easy to compare exam results between different years. It would be more valid to separate students into flipped and non-flipped groups and run the course in parallel, but I feel that resource constraints would make this infeasible for MFCS.

6 Conclusions

My preconception when starting this survey was that flipping was much hyped but little studied. This turned out to be unfair: a reasonable amount of research has now taken place, mostly over the last five years. The result seems to be that flipping is both effective at creating learning, and popular amongst students. Of course, many questions and concerns remain – I have sketched several above. In particular, I hope that future studies will make clear the workload impact of flipping over the long term. Nonetheless, there seems good cause to begin integrating flipping into STEM education programmes, at least in a prototype form.

More broadly, this review has emphasised to me the difficulty in coming to firm conclusions on whether teaching techniques are effective. Technological change is a constant in education practice. Our dilemma is not whether to embrace or resist it change, but how to best take advantage of it. It is important that we make these decisions after reflection, using the best available evidence.

References

[1] Mathematical Foundations of Computer Science, `http://www-course.cs.york.ac.uk/mfcs/`.

[2] Designing and Maintaining Software, `http://www-course.cs.york.ac.uk/dams/`.

[3] Education Resources Information Center (ERIC). `http://eric.ed.gov/`.

[4] J. Bergmann and A. Sams. *Flip Your Classroom.* International Society for Technology in Education, 2012.

[5] J. L. Bishop and M. A. Verleger. The flipped classroom: A survey of the research. In *Proc. 120th ASEE Annual Conference and Exposition*, 2013.

[6] K. R. Clark. The effects of the flipped model of instruction on student engagement and performance in the secondary mathematics classroom. *Journal of Educators Online*, 2015.

[7] R. S. Davies, D. L. Dean, and N. Ball. flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. *Education Technology Research and Development*, 2013.

[8] J. A. Day and J. D. Foley. Evaluating a web lecture intervention in a human computer interaction course. *IEEE Transactions on Education*, 2006.

[9] F. Dochy, M. Segers, P. V. den Bossche, and D. Gijbels. Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, 2003.

[10] J. M. Fautch. the flipped classroom for teaching organic chemistry in small classes: is it effective? *Chemistry Education Research and Practice*, 2015.

[11] P.-E. Gobry. What is the flipped classroom model and why is it amazing? *Forbes*, November 2012. `http://www.forbes.com/sites/pascalemmanuelgobry/2012/12/11/what-is-the-flipped-classroom-model-and-why-is-it-amazing-with-infographic/`.

[12] J. L. Jensen, T. A. Kummer, and P. D. d M Godoy. improvements from a flipped classroom may simply be the fruits of active learning. *Life Sciences Education*, 2015.

[13] M. Kettle. Flipped physics. *Physics Education*, 2013.

[14] B. Kuykendall, M. Janvier, I. Kempton, and D. Brown. Interactive whiteboard technology: Promise and reality. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 2012.

[15] B. Love, A. Hodge, N. Grandgenett, and A. W. Swift. Student learning and perceptions in a flipped linear algebra course. *International Journal of Mathematical Education in Science and Technology*, 2013.

[16] J. McGivney-Burelle and F. Xue. Flipping calculus. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies*, 2013.

[17] M. Moravec, A. Williams, N. Aguilar-Roca, and D. K. O'Dowd. Learn before lecture: A strategy that improves learning outcomes in a large introductory biology class. *CBE Life Sciences Education*, 2010.

[18] T. Rosenberg. In flipped classrooms, a method for mastery. *New York Times*, October 2013. `http://opinionator.blogs.nytimes.com/2013/10/23/in-flipped-classrooms-a-method-for-mastery/`.

[19] M. Touchton. Flipping the classroom and student performance in advanced statistics: evidence in a quasi-experiment. *Journal of Political Science Education*, 2015.

[20] J. D. Tune, M. Sturek, and D. P. Basile. flipped classroom model improves graduate student performance in cardiovascular, respiritory, and renal physiology. *Advances in Physiology Education*, 2013.

[21] L. S. Vygotsky. *Mind and Society: the development of higher mental processes*. Harvard University Press, 1978.

[22] S. G. Wilson. The flipped class: a method to address the challenges of an undergraduate statistics course. *Teaching of Psychology*, 2013.

[23] J. R. Winquist and K. A. Carlson. Flipped statistics class results: better performance than lecture over one year later. *journal of statistics education*, 2014.