12

Can the flipped classroom enhance student learning?

Melanie Ganz-Benjaminsen

Department of Computer Science
University of Copenhagen

Introduction

The flipped classroom (FC) is an interesting teaching tool and its users claim they can engage their students more, increase the students' learning opportunities and get instant feedback on their own teaching at the same time. An in depth treatment on the rise of FC as well as the theoretical background is given in Bishop et al., 2013. The authors also highlight the key concept behind FC as in being pre-recorded lectures or reading material that can be assigned to students as homework, leaving class time open for interactive learning activities—activities that cannot be automated or computerized. Specifically, Bishop et al., 2013, define FC as interactive group learning activities inside the classroom, and direct computer-based individual instruction outside the classroom. While Bishop et al., 2013, in their analysis strictly exclude designs that do not employ videos as an outside of the classroom activity, in general assigning reading outside of class is commonly employed in FC and will be employed in our study as well.

A front figure for FC at the University of Copenhagen is Professor Jan H. Jensen from the department of chemistry (Jensen, 2014). Professor Jensen's approach (Jensen, 2015) lets the students acquire knowledge before they come to class by watching videos or reading a book and checks the knowledge they have acquired by performing a quiz at home. The lecture then serves as a question and answer session where he discusses the issues/points that most of the students missed. He picks up a common question/misconception and starts clarifying the issue but first lets the students try to understand the issue by discussing it with their neighbors (peer instruction). He then asks questions via a student-response system (SRS). The SRS can be used to ask the students a question, to receive their answers in real time by smartphones, tablet or web browsers and see their responses live in class. After students have discussed for a short while, he lets them vote on the issue to check if the knowledge was acquired correctly. If more than, 75% of the students get the answer correct, he moves on to the next issue. If less than 75% but more than 40% give the correct answer, he discusses further and lets them vote again. If less than 40% of the students get the answer right, he then gives an additional short lecture on the topic to ensure the correct understanding of the issue. According to Professor Jensen's experience (Jensen, 2014) this type of teaching format and the use of SRS can counteract the one-way communication of traditional lectures and provide a break every now and then, which is necessary to maintain student concentration. Furthermore, SRS can stimulate class discussions and peer-to-peer instructions as the students become active participants in their own learning. A nice feature is that the teaching format also works well for large classes, but of course the Q&A session needs to be better prepared if the class is larger.

In this work, FC is applied in the course Medical Image Analysis. A few lectures are changed from "knowledge-dissemination" lectures to FC lectures that support a more active learning approach. The goal of this is to check if student learning is enhanced after the use of FC. This is evaluated by two quantitative quizzes before and after the lectures as well as a qualitative interview afterwards.

Background

The course used for this experiment is Medical Image Analysis (MIA, 2017) at the Department of Computer Science of the University of Copenhagen. The course is a 7,5 ECTS course running in block 1 and one of the restricted elective courses in the Master of Science (MSc) program in Computer Science. Furthermore, the course is also an elective course in the MSc in Physics program for the study area Medical Physics. The teaching format mentioned in the course description consists of lectures, exercises, and assignments. In practice, this results in two lectures (Monday and Friday morning) of about 2 hours each accompanied by one afternoon session (Tuesday) with a teacher present, where the students can get help with the assignments. The expectation that is communicated at the start of the course is that the students should read the relevant parts of the course book before the lecture, and hence the "lectures" should be quite interactive. Unfortunately, even though the intention of the course is to be very interactive, the mere request to have read the relevant chapters in the book before the lecture, does not lead the students to do this. This in turn often leads to a one-way lecture that might sometimes be interspersed with study group work, but where the students mostly only sit and listen. While the course in general receives good evaluations, it seems the students lack the ability to transfer the learned knowledge to a new context such as project work after the course. This is a problem for the master students attending the course, since they should be able to transfer what they have learned in the course to research or Master projects that they conduct in the department of computer science.

In the interest of congruency and to avoid confusing the students, FC is tested in two lectures. The first is a short (20 minute) expert lecture on a freely chosen topic by the lecturer. The second lecture is a full lecture (2 hours) and covers the topic of basic medical statistics. This lecture is largely independent of the other content covered in the course and therefore offers the possibility to test FC in an isolated fashion.

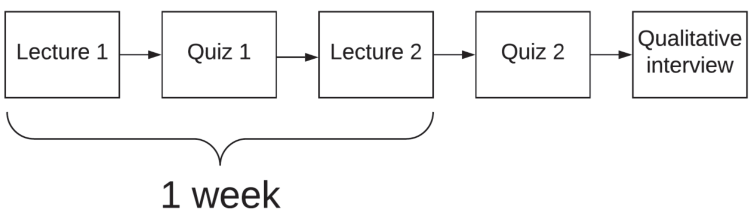

Fig. 12.1. An overview of the experimental setup. Lecture 1 and 2 were given one week apart. In between, the students had to fill out a quiz on the Absalon platform that checked their knowledge about basic statistics. After the second lecture, they were asked to take the same quiz again. Finally, a student was interviewed regarding her experience of the flipped classroom approach.

Methods

The experimental setup to test if FC improves student learning was planned and discussed together with Camilla Østerberg beforehand. The study setup consisted of two lectures interspersed with quantitative quizzes and completed by a qualitative interview. An overview of the experimental setup is shown in Figure 12.1.

The first lecture introduced the students to the lecturer and to a simplified version of FC. The topic of the lecture was an aggression project carried out at the Neurobiology Research Unit at Rigshospitalet (www.nru.dk). The students received a scientific article (da Cunha-Bang et al., 2017) describing the project as well as a link to a documentary about the project by Danmarks Radio [Vold på hjernen] beforehand. Small quizzes and feedback session were used throughout the lecture in order to introduce the students to a flipped classroom type of teaching approach as well as to acquaint the students with the student-response system (SRS) used in the lectures, Shakespeak (https://www.shakespeak.com/). The first lecture was also used to get demographic information about the students.

After the first lecture, the students were assigned a quiz via Absalon that checked their knowledge about basic statistical tests. The quiz consisted of ten multiple-choice questions. A full list of all questions and answers can be found in the supplementary material. There was no time limit on the quiz, the correct answers were not accessible to the students afterwards and they had only one attempt at performing the quiz. The questions were shown one at a time and shuffling of the questions not possible. The students had to hand-in the quiz five days after the first lecture and two days before the second lecture. Additionally, the students were given literature (Van Emden, 2012) to refresh their knowledge if they came upon areas of the quiz where they felt a need to do so.

The second lecture, given one week after the first, introduced the students to basic statistical tests used in medical imaging. The lecture consisted of a very short introduction of theory and then a series of example problems where the students had to vote which statistical test would be appropriate for the example problem. After each small quiz, the example problems were discussed and solved together in class. The second half of the lecture consisted of interactive group work in groups of four to five students where the students designed flowcharts guiding them to the appropriate statistical test given some example data. Finally, the students applied their flowcharts to a final problem that had been introduced at the very beginning of the lecture.

After the second lecture, the students had to complete another quiz, which was identical with the pre-lecture quiz and again checked their knowledge.

Statistical significance of improvement between quiz 1 and 2 was tested with a non-parametric one-sided (left-tailed) Mann Whitney U-test due to a lack of normality of the scores.

Results

Basic demographic information about the students was collected in the first lecture. Based on the student population from previous years, the students were asked to identify their main field of study between computer science, math, physics or other fields. In the current class the students main field of studies were computer science (24%) and physics (4%) as well as other fields (72%). The students from the other field were mainly students in bioinformatics. The students were also asked to identify their seniority and the majority of students was in the second year of their Master of Science education (83.3%) whereas a few first year (8.3%) and bachelor students comprised the rest of the class (8.3%).

Next, the quantitative results based on the first and second ten-question quiz on Absalon will be described. Thirty-three students filled out the pre-lecture quiz on Absalon and all of them completed it before the flipped classroom lecture. The average student performance was a score of 7.76 ± 0.29 out of 10. Thirty-four students completed the post-lecture quiz on Absalon. Three students completed it right after the lecture, while the majority of class submitted the second quiz within the following days. The average student performance on the second quiz was 8.36 ± 0.27 out of 10.

The quiz can further be analyzed in two fashions. Either looking at the individual students or looking at the individual questions in the pre- and post-lecture versions of the quiz. Thirty students filled out both the pre- and post-lecture quizzes. When comparing the individual students' scores pre- and post-lecture, they show a significant improvement after the lecture ($p<0.05$). Means and standard deviations of the two quizzes as well as the outcome of the statistical test are shown in Table 12.1. When examining the ten different items in the quiz, the students improved their performance on average by 11% for seven questions, while they declined in performance on average by 5% for three questions. The largest improvement was on question 7 where correct answers increased by 30%. As example, the three questions with the largest improvements are shown in Table 12.2.

Table 12.1. An overview of the student performance in terms of the average student score and its standard deviation in the pre- and post-lecture quiz on Absalon. Statistical significance of improvement between quiz 1 and 2 was assessed with a non-parametric one-sided (left-tailed) Mann Whitney U-test and yielded a significant improvement.

	Quiz 1	**Quiz 2**	**p-value of one-sided Mann Whitney U-test**
Student Performance (mean ± std)	7.77 ± 0.31	8.57 ± 0.27	0.033

Table 12.2. An overview of the three (out of ten) items where the correct student answers improved the most between the pre- and post-lecture quiz.

	Quiz 1 – Correct answers	**Quiz 2 – Correct answers**	**Improvement in correct answers**
”What is an example of a statistical test I could use if I have non-normal distributed data?”	67 %	97 %	30 %
”You are a high school teacher interested in how your students compare to others in college readiness. The students grades are normally distributed around an average that we call mu. You compare their SAT scores to the national average of 500. What of the following statements is true?”	55 %	68 %	13 %
”My friend, Bob, believes that his supermarket's prices are lower than mine. We construct a list of identical items and purchase them at our respective stores every week for two months. We realize that our items' prices are normally distributed. Then Bob wants to know if his hypothesis is supported. What should he use?”	63 %	79 %	16 %

Finally, we will describe the results based on a qualitative interview with a student. Four students were approached for an interview, but only one student could make the time for it. During the interview the student was asked if she preferred the more active type of lesson exhibited in the statistics lecture that involves group work, if she had a better learning experience by being forced to use the mathematical concepts in examples/quizzes right away and finally if she had experienced this type of active learning or FC before. The student answered that she definitely preferred the more active type of lesson. She also liked to use the concepts right away to ensure she understood them and could apply the newly acquired knowledge. The student has had classes using FC before and mentioned that she felt better suited to this type of learning than passive lectures.

Discussion

While the quantitative (the two quizzes) as well as the qualitative (the interview) results indicate that student learning improved, this needs to be interpreted cautiously. The most severe limitation of drawing strong conclusions is in the single arm design of our study. A randomized design as e.g. presented in Wozny, 2018, where half the students experience a traditional lecture, while the other half are exposed to FC would clearly have been preferable. This was unfortunately not possible due to the small number of students in this course (approx. 30). In general, the small sample size is also a limitation for the validity of our statistical finding. While we try to mitigate the small sample size by using non-parametric tests, a larger sample would allow stronger conclusions.

Furthermore, it is not clear if the positive learning effect and students positive perception stems directly from FC or simply from an active learning approach. In a recent review of the literature across higher education O'Flaherty et al., 2015, noted that much of the emerging evidence of improved academic performance might be indirectly related to FC. For example, Jensen et al., 2015, compared a flipped and non-flipped approach and found no significant difference between either approaches when both groups employed active learning techniques in class. Again, our small study cannot address this issue.

Indirectly, there are though some clear benefits of FC that became apparent in our study. First, since the class size was small and the number of students known, the lecturer could ensure that all students participate in the lecture quizzes by simply waiting for the same number of votes to come in as there were students present. This ”forces” the students to participate. Even though this doesn't ensure that they participate actively, it makes it more likely that they do. Second, the group work in the second half of the statistics lecture engaged the students in a different fashion. They not only had to answer pre-defined questions, but also had to design a flow chart themselves. This allowed them to be creative and think about the topic at hand in an alternative fashion which should also support knowledge retention. Finally, after both lectures five students approached me and asked for project or thesis topic work. By now it is clear that two students will begin a master project under my supervision in spring 2019. Two master student projects from a mere contact of two lectures seems also to show that the lectures were interesting. Whether the students interest in project work was inspired by the topic of the lectures or the lecture style is unknown.

Conclusion

The experience with testing a flipped classroom approach was thoroughly positive and is encouraging me to re-design other parts of the course to support either FC or at the least an active learning approach.

For future work, other lecturers at computer science will be presented with our results to stimulate them to try out FC in their courses, since many of the courses at the department could be re-designed to follow an active learning approach. Overall, I gained a lot of experience with modern teaching methods during the course on teaching in higher education and this experiment and am hoping to apply that knowledge in my future teaching endeavors.

Acknowledgements

The author is grateful to Camilla Østerberg, Pieter Jan Kerstens and Martin Nørgaard for helpful hints and valuable suggestions.

References

Bishop, J. L., & Verleger, M. A. (2013, June). The flipped classroom: A survey of the research. In *ASEE national conference proceedings, Atlanta, GA* (Vol. 30, No. 9, pp. 1-18).

da Cunha-Bang, S., Hjordt, L. V., Perfalk, E., Beliveau, V., Bock, C., Lehel, S., ... & Knudsen, G. M. (2017). Serotonin 1B receptor binding is associated with trait anger and level of psychopathy in violent offenders. *Biological psychiatry, 82*(4), 267-274.

Jensen, J.H. (2014). I lecture no more. Keynote speech at “Make a difference - Teach and learn with technology 2014”, Copenhagen 12 November 2014.

Jensen, J. H. (2015). *My flipped classroom: what I did and how I did it* (No. e1546). PeerJ PrePrints.

Jensen, J. L., Kummer, T. A., & Godoy, P. D. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE—Life Sciences Education, 14*(1), ar5.

O’Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The internet and higher education, 25*, 85-95.

MIA (2017). Medical Image Analysis, NDAK10005U, http://kurser.ku.dk/course/ndak10005u.

Van Emden, H. F. (2012). Statistics for terrified biologists. John Wiley & Sons.

Vold på hjernen (2017), https://www.dr.dk/tv/se/kriminel-hjerne/vold-pa-hjernen/vold-pa-hjernen

Wozny, N., Balser, C., & Ives, D. (2018). Evaluating the flipped classroom: A randomized controlled trial. *The Journal of Economic Education, 49*(2), 115-129.

A

Supplementary material to “Can the flipped classroom enhance student learning?”

Melanie Ganz-Benjaminsen

Department of Computer Science, University of Copenhagen

Quiz questions

1) Statistical hypothesis tests can be used to…

distinguish real effects from random variation	True
make sure my paper gets published	False
prove my hypothesis is true	False

2) Statistical significance, or the probability of finding statistical significance is also known as

p-value	True
standard deviation	False
degrees of freedom	False

3) What is the name of a common statistical test?

t-test	True
q-test	False
r-test	False
s-test	False

4) A t-test is a significance test that assesses

The means of two independent groups	True
The medians of two independent groups	False
The standard deviation of three independent variables	False
The modes of two independent variables	False

5) What should be true of your data before using a t-test?

normally distributed	True
skewed	False
two-tailed	False
uniformly distributed	False

6) What is a common type of t-test?

One-sample t-test	True
Variable t-test	False
Null hypothesis t-test	False

7) What is an example of a statistical test I could use if I have non-normally distributed data?

Mann-Whitney test	True
t-test	False
Delta test	False

8) You are a high school teacher interested in how your students compare to others in college readiness. The students grades are normally distributed around an average that we call mu. You compare their SAT scores to the national average of 500. What of the following statements is true?

Your null hypothesis is mu = 500.	True
Your null hypothesis is mu is not equal to 500.	False
Your alternative hypothesis is mu > 500.	False

9) My friend, Bob, believes that his supermarket's prices are lower than mine. We construct a list of identical items and purchase them at our respective stores every week for two months. We realize that our items' prices are normally distributed. Then Bob wants to know if his hypothesis is supported. What should he use?

Bob has a directional hypothesis and should do a paired t-test.	True
Bob has a directional hypothesis and should do an unpaired t-test.	False
Bob has a non-directional hypothesis and should do an unpaired t-test.	False
Bob has a non-directional hypothesis and should do a paired t-test.	False

10) T-tests and other significance tests are frequently criticized. Overrepresentation of statistical significance in research may result in…

Publication bias	True
Confused students	False
Researcher fatigue	False
Lost funding	False