FLIPPING LARGE UNIVERSITY COURSES: HOW DO STUDENT LEARNING GAINS IMPROVE COMPARED TO LECTURES?

Guillaume Schiltz[1], Gerald Feldman[2], Andreas Vaterlaus[1]

[1]ETH Zurich (Switzerland)
[2]George Washington University, Washington DC (USA)

Abstract

We have conducted a pilot study within a physics lecture class of 370 students at a major Swiss research-intensive university. In a one-year undergraduate physics course, we divided the student cohort into two parallel teaching settings. During one semester, we offered a highly interactive flipped learning environment (SCALE-UP) to one group of 52 students and a reformed lecture to the remaining 318 students. In the second semester, all students were taught in the same lecture setting without a flipped learning alternative. Comparing the performance results of both groups we can draw conclusions on immediate and medium-term learning effects. In addition, we analyzed student feedback in both settings that included data related to class attendance, out-of-class preparation and level of intellectual challenge. In this presentation, we will present our results and draw conclusions on implementing flipped learning in large courses at a research university.

Keywords: physics, SCALE-UP, lecture, study time, exam performance, longitudinal effects

1 INTRODUCTION

Flipped learning, also known as flipped class and inverted learning, has emerged in the early 2000s as a pedagogical approach within the student-centered instructional framework [1]. Paralleled by the growing propagation of technology-based instruction, the definition of flipped learning often includes explicit references to video instruction and other web-based facilities [2]. For our purpose, we opt for a more general definition. We consider flipped learning to be an instructional setting where the major part of content delivery is accomplished outside of the classroom and available class time instead is used for engaging students in collaborative and hands-on activities. This definition shifts the pedagogical focus from self-directed learning, which is moved outside of class, to the more interactive learning opportunities that now can be provided during the contact hours. Hereby, flipped learning is an instrument to maximize classroom time for interactive learning exercises while systematically reducing the lecturing part. In physics education, interactive learning has been identified as one of the major teaching assets for quite a long time and is supported by a large body of research [3,4].

Implementing flipped learning, however, is not obvious and relies on many factors related to the local learning and teaching culture, the existing assessment regulations, the curricular boundary conditions and, most important, on scalability considerations. Flipping a class with 30 students might be considered a feasible task, but flipping a lecture with 300 students turns out to be rather challenging and may potentially require considerable investments, such as room reconfiguration and increased teaching manpower [5]. Before any department or university considers adopting flipped learning in a given local context, it will be necessary to identify possible assets and drawbacks beforehand. For this reason, we have conducted a pilot study within a typical physics lecture class of 370 students at a major Swiss research university. Our aim was to identify and to quantify the main benefits of flipped learning compared to physics lectures, based on the premises of a typical European research university.

The present study was conducted to answer the following two questions:

- What are the students' short-term and medium-term performance gains that can be expected from flipped learning?
- Did students in the flipped learning setting develop a different learning behavior and did their attitudes towards the learning goals differ from those of the lecture students?

Published in:
EDULEARN18 Proceedings
Pages: 2635-2643
Publication year: 2018
ISBN: 978-84-09-02709-5
ISSN: 2340-1117

Published by IATED Academy
Edited by L. Gómez Chova, A. López Martínez, I. Candel Torres

DOI: 10.21125/edulearn.2018.0705

2 METHODOLOGY

2.1 Course description

In a one-year undergraduate physics course, we divided the student cohort into two parallel teaching settings. During one semester, we offered a highly interactive flipped class (SCALE-UP pedagogy [6]) to one group of 52 students and a reformed lecture to the remaining 318 students (Fig 1). In the second semester, all students were taught in the same lecture setting, without a flipped class alternative.

Figure 1. Photos of the lecture hall and of the SCALE-UP classroom.

Apart from content delivery, the lecture class included 40 demonstrations and 37 conceptual clicker questions within a Peer Instruction [7] environment. The latter engaged students interactively and provided immediate feedback to the instructor regarding their level of understanding. Thus, the lecture already included some active-learning elements and it was well aligned to the common framework of reformed lectures in physics [8].

The students in the SCALE-UP setting worked through different activities in small groups of 3-4 students each (Fig. 2). Before each class, students started learning about a topic by doing assigned readings and online exercises via *MasteringPhysics* (www.pearsonmylabandmastering.com). In class, the students performed activities that helped them understand the basic concepts from their home reading. They applied these concepts in hands-on experiments and in collaborative discussions for solving conceptual and numerical problems. Lecturing in SCALE-UP was reduced to a bare minimum.

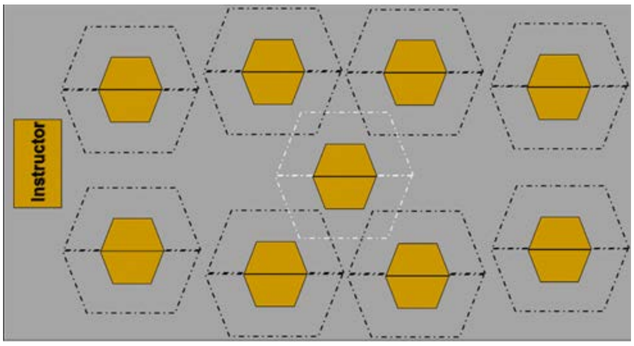

Figure 2. Table arrangement of the SCALE-UP classroom. 9 hexagonal tables (formed with two trapezoids) for 18 groups of three students (two groups per table).

The teaching period for the experimental setting extended over 14 weeks with three weekly contact hours. An additional weekly recitation session (1 hour) allowed students from both settings to discuss numerical problems in groups (15-25 students) together with teaching assistants.

According to university regulations, we were not allowed to use grading as an incentive to control the students' learning behavior or to administer different grading schemes to the separate groups. Grades for both groups were determined by a comprehensive high-stakes final exam that took place eight months after the flipped classroom intervention.

In addition to the performance data, we recorded student feedback data from two separate semester evaluation surveys. These surveys included data related to class attendance, time spent on out-of-class preparation, level of intellectual challenge, and self-confidence in the comprehension of the course material.

2.2 Performance measures

In order to conduct a comparative study of the two different pedagogical settings, we recorded the performance of the complete student cohort (both SCALE-UP and lecture) at different points in time for two different subjects (Fig. 3):

- Physics mid-term exam: 10th week (optional) during the intervention.
- Math (calculus) exam: 3 months after the intervention (as a secondary control variable).
- Physics final exam: 8 months after the intervention.

Both the math and physics final exams are compulsory high-stakes exams. The failing rate for each of those exams is typically higher than 20%. These exams are offered twice a year, and within a given time frame, students can choose the exam session they wish to attend.

Participation in the mid-term exam was optional. As an incentive, however, the result of the mid-term exam could be counted for 10% of the final grade, but only if ameliorating the result of the final exam.

The physics mid-term and final exams included conceptual and numerical questions. In the mid-term exam, 50% of the points could be achieved by conceptual multiple-choice questions, whereas the ratio in the final exam was 40%. Therefore, we were able to split the overall achievement into conceptual and numerical performance components. Conceptual questions assess student understanding of the underlying phenomena rather than the application of the physics material within a mathematical framework. Thus, our study enables us to make a clear distinction between the conceptual understanding and its numerical transfer.

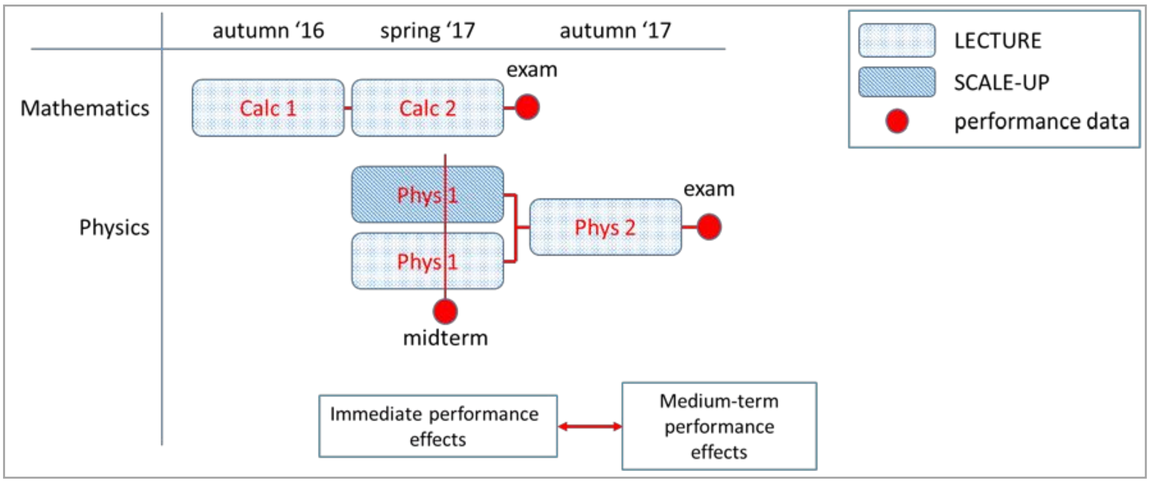

Figure 3. Performance assessment of the student cohort.

Furthermore, the physics final exam was split into one part covering the topics (mechanics) that were introduced during the flipped classroom intervention in spring (Phys1) and another part with the topics (electricity and magnetism) that were covered in autumn without a parallel setting (Phys2). With this distinction, we are able to draw conclusions on longitudinal effects (Phys1) and on how well the learning achievements of the flipped class can be transferred to new topics (Phys2).

2.3 Participants

Related to the complex curricular assessment schedule outlined above and also taking into account dropouts during the course time, we had to deal with a considerable amount of missing data points. Not all students consistently took part in every exam. For instance, 11 of the SCALE-UP students who did not pass the math exam consequently did not take the physics final exam. Throughout the performance analysis, we are only considering students who took part in all assessments, i.e. the math exam and the physics mid-term and final exams. As a result, we had to reduce the overall population to 35 students in the SCALE-UP setting and 133 students in the lecture setting. The data are still sufficient to run statistical tests, even though we have to deal with an unbalanced design.

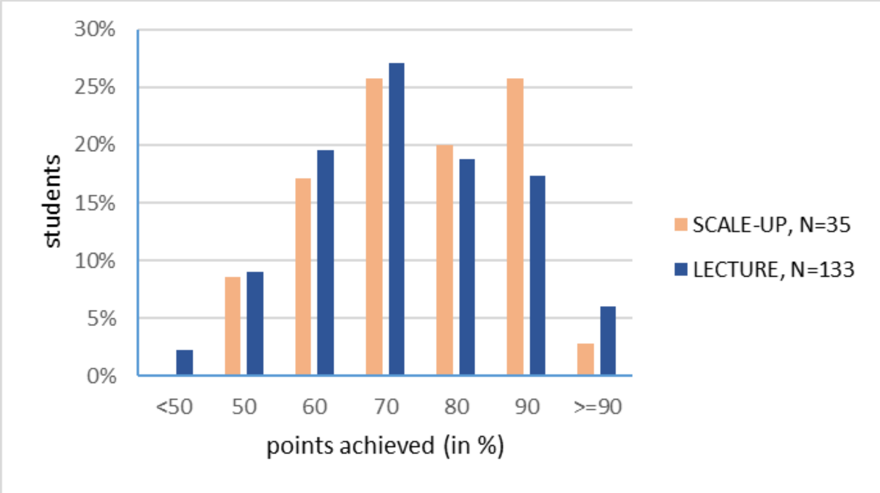

Figure 4. Math performance.

We did not allocate the students randomly to the SCALE-UP class. Instead, we offered the flipped class on a first-come basis. Therefore, it might be argued that students from the lecture and from the SCALE-UP setting may differ with respect to their prior knowledge and that our comparative results may be biased. To rebut this argument, we can consider the students' performance in the math exam. In the past, many studies have shown evidence that achievement in mathematics correlates positively with undergraduate performance in physics (e.g. [9, 10]). A comparison of the math performance from the SCALE-UP and from the lecture students (Fig. 4) shows no significant differences: $t(166)=0.615$, $p=0.539$. For this reason, we assert that our comparative measurements will be unbiased to a certain extent. We came to a similar conclusion based on a conceptual physics pre-test that was given to a sub-sample of students [11].

Finally, the two instructors were expert teachers with extensive experience in their respective teaching settings. In the student survey, questions related to the two instructors resulted in similar positive appraisal.

Table 1. Test statistics of the student performance for each set of questions.

	SCALE-UP Mean (SD)	LECTURE Mean (SD)	t-statistics	Effect size
complete				
Mid-term	61.75 (22.59)	48.99 (20.01)	**3.27 (p=.001)**	0.621
Phys1	50.71 (9.34)	46.66 (11.71)	**2.16 (p=.034)**	0.360
Phys2	46.47 (12.61)	44.75 (11.56)	0.77 (p=.443)	0.146
conceptual				
Mid-term	64.84 (23.57)	48.12 (21.11)	**4.07 (p<.001)**	0.773
Phys1	57.86 (20.01)	46.76 (18.58)	**3.09 (p=.002)**	0.588
Phys2	61.43 (22.96)	62.26 (19.05)	-0.22 (p=.825)	-0.042
numerical				
Mid-term	58.65 (25.00)	49.85 (23.79)	1.92 (p=.056)	0.366
Phys1	68.14 (11.32)	65.28 (16.33)	1.20 (p=.233)	0.185
Phys2	59.07 (17.53)	55.64 (16.83)	1.06 (p=.289)	0.202

$N_{SCALE\text{-}UP}=35$; $N_{LECTURE}=133$; means are referring to the students' performance in percent; SD = standard deviation; effect sizes associated with independent t-tests are computed using Cohen's d. Statistically significant results are marked in bold.

3 RESULTS

3.1 Performance results

For each set of questions (complete, conceptual, numerical), the points achieved in the three physics assessments have been normalized to a percentage scale. In order to compare the performance of students from the SCALE-UP setting to those of the lecture, we have conducted a series of t-tests. The t-test is a common statistical analysis to determine whether the mean of a population significantly differs from the mean of another population. Tab. 1 shows the results from this analysis.

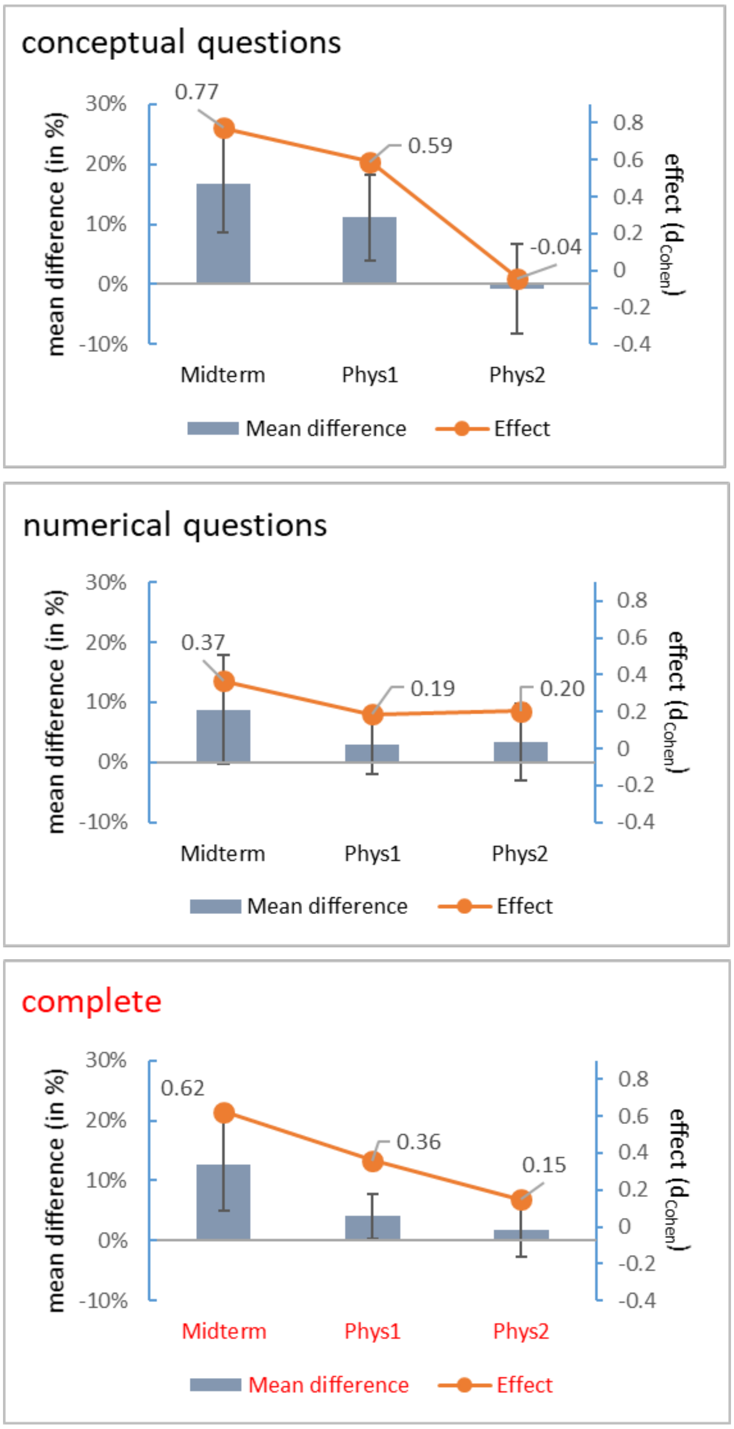

Figure 5. Performance gains of the SCALE-UP students. The gain is calculated by the difference in the means G = $M_{SCALE-UP}$ – $M_{LECTURE}$. Error bars correspond to the 95% confidence intervals. All values are taken from Tab. 1.

In addition to the t-test statistics, we have calculated the corresponding effect sizes. The effect size given by Cohen's d measures the magnitude of mean differences and gives a concrete sense of whether a difference is meaningfully large, independent of whether the difference is statistically significant. Effect sizes of d=0.2 are considered to be small, whereas d=0.5 is related to a medium effect and d=0.8 to a large effect.

With the data from Tab. 1, we have calculated the performance gains of the SCALE-UP students by taking the difference of the SCALE-UP mean value and the corresponding lecture mean value. The resulting gains are plotted in Fig. 5.

3.1.1 Immediate performance effects

The mid-term exam took place during the intervention. Therefore, the mid-term is an adequate instrument to measure immediate effects of the flipped SCALE-UP setting. Fig. 5 clearly shows that the SCALE-UP students outperformed the lecture students in the conceptual questions, whereas their gain in the numerical part is only marginal.

Thus, we see that performance effects attributed to flipped learning are immediately visible. They positively affect the conceptual understanding and do not compromise the skills of solving numerical problems. Those results are confirmed to a certain extent by other independent studies from physics education research (e.g. [12, 13]).

3.1.2 Medium-term performance effects

The Phys1 part of the final exam covered the same topics that were taught during the intervention, but that exam took place eight months later. While comparing the performance related to Phys1 in Fig. 5, we see that the immediate effects still pertain, but that they have diminished considerably.

We can directly compare the performance recorded in the mid-term exam to the performance in Phys1 by running a series of dependent t-tests (for paired samples). The corresponding results are shown in Tab. 2. Both groups have considerably improved their performance in the numerical part. This gain can partially be attributed to exam preparation, where students primarily worked on numerical problems. In the conceptual part, however, the SCALE-UP students significantly reduced their advantage: -6.98% within a 95% confidence interval of [-0.14, -13.82] (*t(34)=-2.07, p=0.046*). Students from the lecture setting had no significant performance changes on the conceptual part.

We can conclude that at a medium timescale the effects of flipped learning are still visible, but have substantially decreased. Unfortunately, comparative studies on medium-term effects are still missing in the literature and we are not able to relate our results to other findings.

Table 2. Test statistics comparing the performance between Mid-term and Phys1.

	Mean (SD/SE)	t-statistics	r
LECTURE			
complete	-2.23 (17.19/1.49)	-1.56 (p=.121)	**0.516**
conceptual	-1.36 (22.70/1.97)	-0.69 (p=.490)	**0.351**
numerical	15.43 (21.84/1.89)	**8.15 (p<.001)**	**0.458**
SCALE-UP			
complete	-11.03 (17.31/2.93)	**-3.77 (p=.001)**	**0.705**
conceptual	-6.98 (19.91/3.37)	**-2.07 (p=.046)**	**0.593**
numerical	9.49 (22.10/3.73)	**2.54 (p=.016)**	**0.468**

$N_{SCALE\text{-}UP}$=35; $N_{LECTURE}$=133; Mean is referring to the paired mean differences: $M_{Phys1} - M_{Midterm}$; SD = standard deviation; SE = standard error; r = Pearson correlation coefficient. Statistically significant results are marked in bold.

3.1.3 Transfer of learning skills

In contrast to the lecture students, the SCALE-UP group had to prepare the topics prior to coming to class. This flipped approach was new to all of the students. Furthermore, the in-class activities of the SCALE-UP group were designed with the goal of training the students in collaborative learning skills. We know that for their class preparations and exam studying, a great number of students were making use of learning groups, and we wondered if the students would naturally adopt those skills without any supervision. In general, we were interested in exploring the question of whether or not the SCALE-UP students were able to transfer those new learning approaches and skills to other subjects outside of a flipped classroom environment.

The topics covered by the Phys2 part of the final exam were taught in a lecture environment, and the corresponding results may give some evidence about the extent of transferability. Unfortunately, the corresponding performance results (Tab. 1 and Fig. 5) show no significant differences between the two groups. We may thus conclude that students in the flipped learning environment were not able to adopt or to transfer their learning skills in a successful way beyond the SCALE-UP setting. Indeed, with only 14 weeks of flipped learning and only 3 contact hours per week, the period may have been too short for a fundamental shift in learning behaviors. Other studies suggest that students require more than a semester to adapt to the new learning method [14].

3.2 Evaluation results

To support and enhance the quality of teaching, ETH Zurich has implemented a sophisticated survey instrument [15]. For each semester, students are invited to evaluate their courses via standardized online questionnaires. The survey switches every year between course evaluations, distributed at the end of the teaching period, and exam evaluations that are made available right after the exams but before the grades have been communicated. The response rate is typically around 40%.

Our survey sets included 20 (course) resp. 24 (exam) responses for the SCALE-UP class and 105 (course) resp. 131 (exam) responses for the lecture. Some of the survey questions explicitly address the learning behavior and the level of intellectual challenge experienced by the students. By analyzing the survey results related to these questions, we are able to draw conclusions about class attendance, time spent on out-of-class preparation, and self-confidence.

Table 3. Selected results from the two student evaluation surveys.

Survey question		Scale	Medians	U-test
A1	How often did you attend the course unit?	1: always 2: usually 3: half of time	1/1	p=.140
A2	How many hours on average per week did you spend doing work outside of contact hours?	1: 0-1h 2: 2-4h 3: 5-7h	2/2	p=.917
A3	I am able to explain the most important material learned in this course to a younger student.	1: not true - 5: abs. true	3.5/3	**p=.004**
B1	I am able to explain the most important material learned in this course to a younger student.	1: not true - 5: abs. true	4/4	p=.216
B2	How many days did you spend for the preparation of the exam?	1: 1-2d 2: 7-13d 3:14-20d 4: >20d	2/2	p=.490
B3	How would you estimate your grade in the examination?	1: failed - 6: full points	3/2	**p=.012**

Ax = course survey, Bx = exam survey. Medians: $Mdn_{SCALE-UP}$ / $Mdn_{LECTURE}$. Statistically significant results are marked in bold.

Tab. 3 shows the questions with the corresponding median values. For each of the questions we have conducted a Mann-Whitney U-test. This statistical analysis is testing whether two samples (here SCALE-UP and LECTURE) are likely to derive from the same population. If the test yields a significant result, the two samples are considered to be different.

Course attendance was not compulsory and the high degree of attendance (A1) mirrors the students' high engagement in both settings.

Two interesting outcomes are apparent. First, students from both groups spent about the same amount of time for their out-of-class studying (A2) and for their exam preparation (B2). Even though this finding is not completely new [16], the question of time investment is often an issue during discussions about flipped learning. It is worth mentioning that we did not enforce any pre-class preparation by offering incentives. Other studies have suggested that incentives are imperative for a functioning flipped learning environment [17].

The second surprising finding concerns the questions related to students' self-confidence (A3, B1, B3). In the course survey, the SCALE-UP students exhibited a significantly higher degree of self-confidence. However, this difference was leveled out in the exam survey, whereas here the grade expectation, a second confidence indicator, now was significantly higher for the SCALE-UP students.

4 CONCLUSIONS

Coming back to the main research questions of this study, we can summarize that:

- During the intervention period, students from the flipped SCALE-UP group outperformed students from the lecture setting. This performance gain, however, was substantially reduced when evaluated over the medium-term scale.
- For those students who participated in the 14-week flipped SCALE-UP group, we could not identify any transfer or modification of learning behavior that would induce better performance outside of a dedicated flipped learning setting.
- Compared to the lecture students, students from the flipped SCALE-UP group did not invest more overall study time, even though they had to come prepared to class.
- The SCALE-UP students manifested an increased level of self-confidence in their own learning achievements.

Our study has some limitations, being based on a single intervention with only one cohort of students. However, we carried out the study within our local curricular, time, legal and resource constraints. Those constraints cover the typical setting of a European research-intensive university, and we hope that policy-makers and teachers can profit from our results, when they are confronted with making decisions and weighing the relative benefits of a shift to flipped learning.

In the more general context of flipped learning research, we were able to make a contribution to some important aspects such as transferability or longitudinal effects that are still missing the literature [18].

REFERENCES

[1] Flipped Learning Network (FLN). *The Four Pillars of F-L-I-P™*, 2014. Retrieved from https://flippedlearning.org/

[2] L. Johnson, et al., *NMC Horizon Report: 2015 Higher Education Edition.* Austin/TX: The New Media Consortium, 2015. Retrieved from http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf

[3] J.M. Fraser, et al., "Teaching and physics education research: bridging the gap," *Reports on Progress in Physics,* vol. 77, no. 3, 032401, 2014.

[4] C. Crouch, E. Mazur, "Peer Instruction: Ten Years of Experience and Results," *Am. J. Phys.*, vol. 69, pp. 970-977, 2001.

[5] E. Brewe, et al., "Costs of success: Financial implications of implementation of active learning in introductory physics courses for students and administrators," *Phys. Rev. Phys. Educ. Res.,* vol. 14, 010109, 2018.

[6] R.J. Beichner, et al. "Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) project." in *PER-Based Reform in University Physics*, vol. 1. (E.F. Redish, P.J. Cooney eds.), College Park/MD: American Association of Physics Teachers, pp. 1-42, 2007.

[7] E. Mazur, *Peer instruction: a user's manual.* Upper Saddle River/NJ: Prentice Hall, 1997.

[8] S. McKagan, et al., "Reforming a large lecture modern physics course for engineering majors using a PER-based design," *AIP Conf. Proc*., vol. 883, pp. 34-37, 2007.

[9] H.T. Hudson, R.M. Rottmann, "Correlation between performance in physics and prior mathematics knowledge," *J. Res. Sci. Teach.,* vol. 18, pp. 291–294, 1981.

[10] D.E. Meltzer, "The relationship between mathematics preparation and conceptual learning gains in physics: a possible 'hidden variable' in diagnostic pretest scores," *Am. J. Phys.*, vol. 70, pp. 1259–1268, 2002.

[11] G. Schiltz, et al., "Active-Learning settings and physics lectures: A performance analysis," *Proceedings of the GIREP-ICPE-EPEC Conference 2017*, Dublin, in press.

[12] C. Hoellwarth, et al., "A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms," *Am. J. Phys.*, vol. 73, pp. 459–462, 2005.

[13] M.A. McDaniel, et al., "Dissociative conceptual and quantitative problem solving outcomes across interactive engagement and traditional format introductory physics," *Phys. Rev. Phys. Educ. Res.*, vol. 12, 020141, 2016.

[14] A. Roehl, et al., "The flipped classroom: An opportunity to engage Millennial students through active learning strategies," *Journal of Family and Consumer Sciences*, vol. 105, no. 2, pp. 44–49, 2013.

[15] www.ethz.ch/services/en/teaching/academic-support/teaching-evaluation.html

[16] W. He, et al., "The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions," *Learning and Instruction*, vol. 45, pp. 61-71, 2016.

[17] D.J. Harrison, et al., "Assessing the effectiveness of a hybrid-flipped model of learning on fluid mechanics instruction: overall course performance, homework, and far- and near-transfer of learning," *European Journal of Engineering Education*, vol. 42, no. 6, pp. 712-728, 2017.

[18] A. Karabulut, et al., "A systematic review of research on the flipped learning method in engineering education: Flipped Learning in Engineering Education," *British Journal of Educational Technology*, vol. 49, no. 3, pp. 398–411, 2017.