Collaborative Learning vs. Lecture for Introductory Physics Students: A One-Year Longitudinal Study

G Feldman\textsuperscript{1,2}, G Schiltz\textsuperscript{2} and A Vaterlaus\textsuperscript{2}

\textsuperscript{1}George Washington University, Washington, DC, USA
\textsuperscript{2}ETH Zürich, Switzerland

E-mail: feldman@gwu.edu

Abstract. A collaborative group-learning classroom was implemented on a trial basis at ETH Zürich in the Spring 2017 semester. Data on conceptual understanding, problem-solving ability, and student feedback were collected from the class and compared to a parallel large lecture class. In addition, a comprehensive Final Exam was given in January 2018 to assess the overall learning achievements of the students in both sections. These comparative data will be presented in order to gauge the effectiveness of the collaborative pedagogical approach.

1. Introduction

Studies of undergraduate STEM education demonstrate that students need to be actively engaged in the learning process in order for it to be effective. A passive lecture environment (“teaching by telling”) has been shown to be largely ineffective in developing students’ skills in critical thinking and problem solving [1]. Despite abundant data indicating the shortcomings of the conventional lecture format, however, it continues to be the prevalent instructional mode in most institutions.

Collaborative group-learning approaches such as “studio physics” [2] or SCALE-UP [3,4] have gained popularity in the United States – in this classroom format, students work together in small groups on conceptual and numerical problems, while the instructor serves as a facilitator or “coach” instead of a lecturer. By merging this collaborative methodology with the integration of a broad variety of pedagogical activities (including conceptual/numerical exercises, laboratory experiments, and possibly computer simulations), a dynamic collective learning environment is created which fully engages students. In the Spring 2017 semester, we embarked on an implementation of SCALE-UP as a pilot project in an introductory physics class at ETH Zürich.

2. Classroom Setup

The Physics 1 class at ETH Zürich consisted of approximately 370 students, in a range of disciplines including environmental science, food science, earth science and agricultural science. We offered the SCALE-UP collaborative class to 54 students on a voluntary enrollment basis. Surprisingly, at the time of registration, there were 94 students interested in signing up – thus, we found ourselves with a waiting list of 40 students. However, we could only accept 54, so the remaining 320 students were assigned to the parallel lecture class that was compared with the SCALE-UP class.

The physical configuration of the classroom consisted of 9 hexagonal tables (each formed with two trapezoidal tables) – the 54 students were organized into 18 groups of three (two groups per table).
The initial classroom layout and the reconfigured SCALE-UP format are shown below in Fig. 1. Portable whiteboards (60 cm × 90 cm) were distributed to each group for working together on numerical and conceptual questions in class. An electronic response system (“clickers”) based on TurningPoint [5] was used to deliver conceptual questions to the class, following the Peer Instruction method pioneered by Mazur [6]. Several photos of the classroom itself, as compared to the lecture hall, are depicted in Fig. 2.

Figure 1. Classroom configuration with 9 hexagonal tables for SCALE-UP (right panel) as adapted from a standard classroom layout with movable trapezoidal tables (left panel).

Figure 2. Photos of the SCALE-UP classroom (top left) and the lecture hall (top right). The two photos on the left show students working on their whiteboards, assisted by the instructor and the TA’s. The photo on the bottom right shows a student presenting his group’s results on their whiteboard.

While the large lecture class (with A. Vaterlaus) was delivered in German, the SCALE-UP class (with G. Feldman) was taught in English. This did not pose any problem for the Swiss students, since their English language abilities were uniformly excellent. To facilitate pre-class preparation and post-
class problem-solving practice, we obtained 60 student licenses from Pearson publishers for the MasteringPhysics online system [7], with a link to the Giancoli introductory physics textbook [8]. This served as the primary textbook for the SCALE-UP class, and conceptual and numerical exercises were offered online for the students. The acquisition of these 60 licenses was fully funded by the Department of Physics at ETH, so the students received this benefit for free.

The instructional schedule for both settings consisted of one 3-hr class on Monday morning and a 1-hr recitation session on Thursday with a Teaching Assistant (TA). For the SCALE-UP class, each academic hour was actually 50 minutes followed by a 10-minute break, so the in-class contact time essentially amounted to only 150 mins = 2.5 hrs. The lecture class hour was 45 minutes followed by a 15-minute break. The SCALE-UP class had three TA’s who assisted with the collaborative activities in the classroom on Monday (with G. Feldman) and who ran the recitation sessions (each with about 18 students) which mainly served as question/answer periods. The Physics 1 class did not have an associated laboratory session.

The physics topics covered in both classes were totally aligned and were typical for a first-semester physics course: kinematics, forces, energy, momentum, rigid-body motion, torque, oscillations, fluids, heat and thermodynamics. Students were expected to come to class prepared to work on the group activities together, including conceptual questions that they would discuss and numerical problems that they would solve. Solutions to these questions and problems were reported by the students themselves in the classroom, displaying their group’s whiteboard work to the rest of the class and explaining their solutions. Formal lecture in the conventional sense was reduced to a minimal level, and while there were certainly many engaging discussions about the material (either with the TA’s in the groups as they worked or by the instructor with the entire class as solutions were presented), there was no extensive “transmission” of vast amounts of physics content for extended periods of time. The focus was primarily on the students, letting them wrestle with the material themselves (with real-time guidance by the instructors, as needed) – it was truly a “student-centered” classroom.

3. Pedagogical Methodology
There were several novel features in the SCALE-UP class that were very different from the lecture environment to which the students were accustomed, including collaborative activities with the portable whiteboards, extensive use of “clickers” to answer conceptual questions [5,6], and the availability of the online questions/problems delivered in the MasteringPhysics system [7]. It should be mentioned, however, that clickers were also used in the large lecture class, but to a lesser extent. As research questions, we wanted to explore how the students would respond to these features, as well as other considerations that applied uniquely to the SCALE-UP approach:

- no lectures – very little formal “transmission” of information
- necessity of coming to class prepared to work on exercises
- reactions to new class elements (clickers, whiteboards, online homework, etc.)
- willingness to speak up and report solutions to the entire class
- different (and dynamic!) classroom environment compared to lecture
- longitudinal retention of physics knowledge and skills over an extended period

What did the students actually do in the SCALE-UP class? The in-class activities consisted of “ponderables” (conceptual or numerical exercises to think about) and “tangibles” (short hands-on exercises or demos involving some equipment and some manipulation). Responses to conceptual questions were recorded using TurningPoint clickers [5] and displayed to the class for discussion and resolution. Some graphical or numerical problems were solved collectively by the group members using their whiteboards and then reported to the entire class. Tangibles consisted of short activities such as: (1) dropping a meterstick and catching it as a means of measuring human reaction time via free-fall acceleration, (2) balancing a meterstick on a fulcrum (like a see-saw) with a known mass on one side in order to find the unknown mass of the meterstick, (3) submerging a weight into a beaker of water on a scale to determine if the scale reading changes due to the buoyant force on the weight, and (4) playing with Newton’s Cradle to observe elastic collisions between steel balls.
Outside of class, all the Physics 1 students (SCALE-UP and lecture) had homework problems assigned in the Moodle online system used at ETH. The weekly problem sets consisted of about 4-5 numerical problems that were optional and were intended to be discussed with the TA’s in the recitation sessions. These assignments could be handed in for feedback from the TA, but they were not formally graded (and most students did not hand them in anyway).

For the SCALE-UP class, the use of the MasteringPhysics online system [7] enabled the creation of pre-class “Warmups” which were conceptual questions designed to help the students gauge their own level of understanding before coming to class. Reading the Giancoli textbook in advance (available as an e-text in MasteringPhysics) and working on the Warmups were the primary means of preparation available to the SCALE-UP students. It is important to note that this pre-class work was expected of the students, and their responsibility for preparing before class was made abundantly clear to them. In addition, online homework problems were offered after class through MasteringPhysics to provide more opportunities for numerical problem-solving practice. While the Warmups and homework sets served as a valuable resource for the students during the semester, it must be noted that all of the online exercises in MasteringPhysics were purely optional. And since we only had 60 student licenses for this product, MasteringPhysics was exclusively available to the students in the SCALE-UP class.

To make some determination about the effectiveness of the SCALE-UP pedagogy compared to the lecture format, we collected data (see Table 1) to provide both a quantitative evaluation and some qualitative feedback on the general impact of our intervention. For summative assessments, we gave an optional mid-term exam to both classes in the tenth week of the semester, and we gave a required common Final Exam (for Physics 1+2 combined) in mid-January of 2018. As an incentive to take the mid-term exam, that score could be counted for 10% of the final Physics 1+2 grade, such that it was a “bonus” that could partially compensate for the Final Exam. For a diagnostic conceptual assessment, we gave the Force Concept Inventory (FCI) [9,10] as a pre-test in the first week of class and as a post-test on the last day of class (both were optional). For formative assessments, we gave short (15 minutes) conceptual/numerical quizzes every three weeks in the recitation sessions (which were also optional). Finally, for student feedback, we offered a mid-semester questionnaire for the SCALE-UP class to check how things were going, and there was also a standard ETH course evaluation administered at the end of the semester (for all courses across the institution).

Table 1. Summary of the data collected for our comparison between the SCALE-UP class and the lecture class. It should be noted that the only required element for all students in both sections of the Physics 1 class was the Final Exam in early 2018 (shown in red). Everything else was optional.

| Summative assessment | common mid-term exam (week #10 – May 2017) |
|----------------------|------------------------------------------|
|                      | common Final Exam (January 2018)         |
| Diagnostic conceptual assessment | Force Concept Inventory (pre/post-tests) |
| Formative assessment | 4 short conceptual/numerical quizzes      |
| Feedback and comments | mid-semester questionnaire                |
|                      | standard ETH course evaluation           |

Note that there was no comprehensive exam at the end of the Physics 1 class in Spring 2017. The second half of the course (Physics 2) was taught in the Fall 2017 semester with only a standard lecture section (i.e. without a partner collaborative section). The Final Exam for the entire course was given
in January 2018, at the end of the full-year sequence of Physics 1 and Physics 2. Thus, the Final Exam occurred 7½ months after the end of the Physics 1 course, so the Physics 1 component of that exam constituted a longitudinal assessment of the physics knowledge and problem-solving skills of those students at the end of that time period. It is also potentially interesting to examine the Final Exam results for the Physics 2 component as well. While the Physics 2 class in Fall 2017 was only taught in a lecture format, there is a question as to whether any problem-solving skills that were developed in the Physics 1 collaborative class were transferred to the Physics 2 portion.

4. Results
The mid-term exam was given to both classes in the tenth week of the Spring 2017 semester, and the results are shown in Fig. 3 below. The mid-term consisted of three conceptual questions and three numerical questions (each worth 12 points), for a total of 72 points. In the figure, the scores are binned in intervals of 4 points. The results shown below pertain to all of the students who opted to take the mid-term exam. The average score for the SCALE-UP class was (59.0 ± 3.5)% for N = 42 students and the average score for the lecture class was (45.8 ±1.7)% for N = 162 students. Based on the standard error computed from the respective score distributions, the SCALE-UP class was observed to have performed statistically better than the lecture class.

![Mid-term Exam Results](image)

**Figure 3.** Results of the mid-term exam for the SCALE-UP class (upper panel) and the lecture class (lower panel). The respective averages are shown by the red arrows in each case.

An item analysis of the mid-term exam was conducted to show the specific scores on each of the six exam questions, and these results are shown in Fig. 4. In each of the six questions, the SCALE-UP class had a higher score than the lecture class. It is also interesting to note that the margin is actually somewhat greater for the conceptual questions than for the numerical ones. For conceptual questions, the difference between the two classes varies from 1.7 to 2.6 points, whereas for the numerical
questions, a difference of 0.7 to 1.4 points is observed. A more detailed analysis of the conceptual and numerical components of the mid-term exam is presented in another GIREP paper [11].

Figure 4. Item analysis for the three conceptual questions (1C, 2C, 3C) and the three numerical questions (4N, 5N, 6N) from the mid-term exam. The SCALE-UP class is shown by the orange bars; the lecture class is shown by the green bars. Each question is worth 12 points.

The FCI [9,10] was given to both classes at the beginning and at the end of the semester. This conceptual assessment was discussed in a previous GIREP paper [12], and the reader is referred to that earlier paper for the detailed analysis. Just to restate the conclusion reported earlier, the normalized gain \( \langle g \rangle \), given by the relation \( \langle g \rangle = (\text{post} - \text{pre})/(30 - \text{pre}) \), which characterizes the fraction of missed points on the pre-test that were recovered in the post-test, was shown to be statistically higher for the SCALE-UP class than the lecture class.

The Final Exam for the entire year-long Physics 1+2 course sequence was given in January 2018, and the results of this comprehensive exam are shown in Fig. 5. These results represent all of the students who took the Final Exam, but it should be noted that this sample differs somewhat from the sample of students who took the mid-term exam (which was optional). The separate results for the Physics 1 and Physics 2 parts are shown in the left and middle panels, broken up into conceptual (8 pts total) and numerical (20 pts total) components. The overall exam total is shown in the right panel. It is clear that the SCALE-UP students still managed to score somewhat higher in the Physics 1 part of the exam than their lecture class counterparts, and this observation applies to both the conceptual and the numerical components of the exam. However, that slight advantage has mostly vanished for the Physics 2 part of the exam. Thus, despite the lengthy time interval between their Physics 1 instruction and the Final Exam itself, we see that the SCALE-UP class still outperformed the lecture class in the particular Physics 1 segment of the exam.

In order to focus more specifically on the details that are relevant to our longitudinal study, we compare the relative results of the mid-term exam (from May 2017) and the Final Exam (from January 2018) in Fig. 6 and Fig. 7. However, for this analysis of the two exams, it was necessary to restrict our sample population to only those students who took both the mid-term and the Final Exam.
latter was mandatory for the course, the mid-term was optional, and therefore the sample population was slightly reduced when we imposed the condition of matched sets. For our data sample that is shown in Figs. 6 and 7, we had $N = 35$ for the SCALE-UP class and $N = 133$ for the lecture class.

![Figure 5. Results of the Final Exam for the Physics 1 portion (left panel), Physics 2 portion (middle) and the overall total (right panel). The Physics 1 and 2 portions are also broken up into separate conceptual and numerical scores, where the maximum score is indicated by the horizontal red line. The SCALE-UP class is shown by the green bars; the lecture class is shown by the orange bars.](image)

In Fig. 6, the raw difference between the SCALE-UP class and the lecture class is shown for conceptual (top) and numerical (bottom) problems from the mid-term exam and the separate Physics 1 and 2 components of the Final Exam. For the conceptual questions, we see a significant edge for the SCALE-UP students in the mid-term exam and the Physics 1 part of the Final Exam. The corresponding effect size (as quantified by Cohen’s $d$ coefficient) is also given – this coefficient 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. Typically, effect sizes of $d = 0.2$ are considered to be small, whereas $d = 0.5$ reflects a medium effect and $d = 0.8$ indicates a large effect. We see that the mid-term differences for conceptual questions constitute a large effect ($d = 0.77$), whereas the Physics 1 Final Exam differences show a medium effect ($d = 0.59$). There is no discernible difference for Physics 2. For the numerical questions, there is a suggestion of a possible positive effect on the mid-term exam ($d = 0.37$), but there is either a very small or a negligible effect for both parts of the Final Exam ($d \leq 0.20$).

In Fig. 7, we are comparing relative “gains” between the Final Exam and the mid-term exam for the Physics 1 questions only, for the SCALE-UP class (top) and the lecture class (bottom). Here we see that the SCALE-UP students have lost ground with respect to the concepts, although they have improved somewhat on the numerical aspects. This numerical improvement is actually more pronounced for the lecture class, although there is no gain whatsoever for them on the conceptual side. Given that the two exams were $7\frac{1}{2}$ months apart, it could be that the numerical skills of both student populations improved with greater practice over this time interval, including the enhancement of their
basic skills in Physics 2. While concepts were presented in both sections of Physics 1, there was a greater emphasis on this aspect in the SCALE-UP class. However, it appears that while the SCALE-UP students showed a statistically better performance on conceptual questions in the mid-term, they could not maintain their performance on the Final Exam, hence the negative gain reflected in Fig. 7. On the other hand, the lecture students performed almost the same on conceptual questions in both exams, hence the nearly zero gain.

**Figure 6.** Difference (SCALE-UP – Lecture) for scores on conceptual questions (top) and numerical questions (bottom) for the various exams. The effect size given by Cohen’s $d$ coefficient is also shown.

**Figure 7.** Difference (Final Exam – Mid-term) for scores on conceptual and numerical questions (Physics 1 only) for the SCALE-UP class (top) and the lecture class (bottom). The effect size given by Cohen’s $d$ coefficient is also shown.

**Figure 8.** Student grades for their first-year mathematics course at ETH Zürich. The SCALE-UP class (N = 35) is shown by the orange bars; the lecture class (N = 133) is shown by the blue bars.
One important consideration in the comparison between the SCALE-UP and lecture sections is the question of population equivalency. In our investigation, we did not assign students randomly to the SCALE-UP class – the registration for that class was based entirely on voluntary enrollment. Thus, one could ask the question: Does the SCALE-UP class consist of the self-selected “best” students? To address this question, we were able to make a direct independent comparison between the students in these classes based on their year-long mathematics course during the first year at ETH. We regard this as a reasonable validity test for equivalency – many studies have shown evidence that achievement in mathematics correlates positively with undergraduate performance in physics [13,14]. Utilizing the data from the same restricted set of students presented in Figs. 6 and 7, we show a histogram of their first-year math scores in Fig. 8. It appears that the two student populations are essentially equivalent. We came to a similar conclusion based on the FCI pre-test that was given to the SCALE-UP class and the lecture class at the start of the Spring 2017 semester [12]. So we are confident that we do not have any particular bias in our comparisons between the SCALE-UP class and the lecture class.

5. Conclusions
In summary, we have conducted a collaborative SCALE-UP class at ETH Zürich in an introductory physics course with 54 first-year undergraduate students. We found that the students were highly engaged during the class period, and they seemed to be well prepared for the classroom activities. The students were not at all hesitant to speak up during class – group members freely reported their results to the entire class and follow-up questions flowed naturally from the students during the class period.

A detailed comparative study was performed by collecting data from the SCALE-UP section and from a parallel lecture section. In order to ascertain the equivalency of these two student populations, course grades for their first-year mathematics class were used to verify that the students in both sections were essentially equivalent. In terms of assessments, FCI scores on pre/post-tests showed a higher gain for the SCALE-UP section as compared to the lecture section. Furthermore, results of an identical mid-term exam given to both sections also showed a higher average exam score for the SCALE-UP students. However, in a comprehensive Final Exam given 7½ months later, there were only relatively small differences between the two sections.

We would like to point out several constraints in our study. First, the actual contact time with the students for the SCALE-UP intervention was limited to only 3 hours on one class day per week. One extra hour for the recitation session, on a separate day, did not specifically involve collaborative activities since it was reserved for question/answer periods with the TA. By contrast, the contact time with introductory physics students at George Washington University is typically 5 hours per week. Second, the overall academic schedule for these ETH students is extremely intensive – they are in class for about 30-32 hours per week. This makes it very difficult to expect them to work on homework problems outside of class, simply due to a lack of time. Third, since none of the assignments were required (or actually counted for a grade, except the mid-term exam, which essentially constituted a “bonus”), this contributes even more to a potential lack of incentive for practice outside of class. This issue also pertained to the level of participation in the assessments used for our comparative study – for example, participation in the mid-term exam was highly variable due to the fact that it was purely optional for the students.

Despite these constraints, the collaborative approach was observed to be effective in the near term (i.e. during the period of the intervention), but the effects of this intervention were not long lasting. To make the SCALE-UP class more successful, there is a need to provide time for the skills being developed in the classroom to evolve and “sink in” through a regular sequence of practice at home and feedback by the instructor. However, if this cycle is not followed, then the students are not really able to “close the loop” on the instructional benefits of the collaborative setting. While the students’ performance in the SCALE-UP section was impressive during the intervention period, one could argue that even better results might be obtained if more time could be allocated to in-class activities and homework outside of class, as well as mandatory assignments that keep the students on track as the
semester progresses. In addition, better data for comparative studies could be collected if participation in the assessments would be required for all students, even if those assessments do not count formally for their course grade.

Moving forward, it would be very interesting to try the SCALE-UP collaborative approach in another venue, with some of the improvements mentioned above. While the educational systems of various European countries can be rather different, the basic motivations of the students should be similar, in general, and we have found that the “novelty factor” for this type of collaborative group-learning environment was appealing to the students. Making another pilot study and collecting a robust set of assessment data would prove to be a big step in validating the effectiveness of this approach and confirming its universal applicability.

Acknowledgements

One of us (G.F.) would like to thank the Department of Physics and the Office of the Rektor at ETH for support during the Spring 2017 semester. The warm hospitality of the entire Physics Education Group, including my co-authors and other group members, was deeply appreciated.

References

[1] McDermott L C 2001 Oersted Medal Lecture 2001: Physics Education Research – The Key to Student Learning American Journal of Physics 69 1127 DOI:10.1119/1.1389280
[2] Wilson J M 1994 The CUPLE Physics Studio The Physics Teacher 32 518 DOI:10.1119/1.2344100
[3] Beichner R J et al 2007 The Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) Project Research-Based Reform of University Physics ed E F Redish and P J Cooney (College Park, MD: AAPT)
[4] SCALE-UP web site is located at: http://www.ncsu.edu/PER/scaleup.html/
[5] TurningPoint web site is located at: http://www.turningtechnologies.com/
[6] Mazur E 1997 Peer Instruction: A User's Manual (Upper Saddle River, NJ: Prentice-Hall)
[7] MasteringPhysics web site is located at: https://www.pearsonmylabandmastering.com/
[8] Giancoli D C 2009 Physics for Scientists and Engineers, 4th Edition (Upper Saddle River, NJ: Pearson Prentice-Hall)
[9] Hestenes D, Wells M and Swackhamer G 1992 Force Concept Inventory The Physics Teacher 30 141 DOI:10.1119/1.2343497
[10] Hestenes D and I. Halloun I 1995 Interpreting the Force Concept Inventory The Physics Teacher 33 502 DOI:10.1119/1.2344278
[11] Schlitz G, Feldman G and Vaterlaus A 2017 Active-learning settings and physics lectures: a performance analysis Proc. GIREP-ICPE-EPEC Conf. 2017 (Dublin) Journal of Physics: Conf. Series (in press)
[12] Feldman G, Schlitz G and Vaterlaus A 2017 Collaborative Group Learning in a Swiss Introductory Physics Class Proc. GIREP-ICPE-EPEC Conf. 2017 (Dublin) Journal of Physics: Conf. Series (in press)
[13] Hudson H T and Rottrmann R M 1981 Correlation between Performance in Physics and Prior Mathematics Knowledge Journal of Research in Science Teaching 18 291 DOI:10.1002/tea.3660180403
[14] Meltzer D E 2002 The Relationship between Mathematics Preparation and Conceptual Learning Gains in Physics: A Possible “Hidden Variable” in Diagnostic Pretest Scores American Journal of Physics 70 1259 DOI:10.1119/1.1514215