Abstract

This study assessed the impact of flipped instruction on students' out-of-class study time, exam performance, preference, motivation, and perceptions in two sections of a large undergraduate chemistry course. Flipped instruction caused a shift in student workload without appreciably changing the overall study time. The treatment impact on student performance gradually diminished over time, showing a small but statistically significant effect with the final exam. No marked interaction was identified, indicating that flipped instruction benefited students of diverse backgrounds uniformly. Students in the flipped section **showed mixed feelings with about one fifth of them displaying** polarized attitudes. **Open-ended student survey responses revealed non-compliance with pre-class studying as a serious** implementation issue: **By slowing down the overall pace of the class, it negatively affected students with different study behaviors and characteristics in ways** that partly explained the small, diminishing treatment effect and absence of marked interaction.

Keywords: flipped instruction, chemistry, study time, exam performance, perception

1. Introduction

In recent years, flipped instruction, also known as inverted instruction, has attracted growing attention from both teaching and research communities as a promising instructional technique. Before 2013, we have only identified a small handful of empirical studies in higher education addressing this topic (see a brief review below). The past two years have seen a surge in the number of both case studies and empirical research, as universities worldwide are adopting the flipped pedagogy. The excitement over flipped instruction stems from its capacity to combine purported benefits of online instruction with active learning techniques into a new pedagogy. The essence of the flipped pedagogy is to stage learning of new material before class to free up class time for more in-depth explanation, practice, and productive use of knowledge.

Researchers have not yet reached a consensus on a formal definition of flipped instruction. For early pioneers (Bergmann & Sams, 2008), the intent of flipped instruction was to eliminate lectures in class (i.e., students could watch video podcasts at home before class) so that material that had traditionally been assigned as homework could be completed in class with more student-centered and inquiry-based activities. This idea has led to the general conception of flipped instruction as “events that have traditionally taken place inside the classroom now take place outside the classroom and vice versa” (Lage, Platt, & Treglia, 2000). Bishop and Verleger (2013) suggested including only the studies with computer-based pre-class instruction. We believe qualifying instructional medium is unnecessary and unjustified. It had been shown in a quasi-experimental study with over 800 students that pre-class reading assignments supplemented with worksheets could be as effective as pre-class videos in increasing exam

performance using the flipped pedagogy (Moravec, Williams, Aguilar-Roca, & O'Dowd, 2010).

Based on the discussion above, in this study, we define flipped instruction as having three attributes. Flipped classrooms should feature (a) mandatory pre-class learning of new material followed by (b) in-depth explanation, practice, and productive use of knowledge in class through active learning techniques, where (c) class attendance is mandatory. All three features are necessary. First, pre-class learning is an integral part of instruction. Long before flipped instruction was studied as a distinct pedagogy, instructors were known to assign textbook material for students to read before class. In traditional classrooms, however, pre-class learning was often not enforced and instructors would cover the pre-assigned material in class anyway. In a flipped classroom, pre-class instruction is designated for teaching (factual) knowledge that will not be repeated in class except for brief reviews. Secondly, productive use of knowledge should dominate class time in order to promote conceptual understanding. Traditional classrooms also adopt active learning techniques. Due to a packed schedule, however, active learning often accounts for a small proportion of the class time. In a flipped classroom, since lectures are, to a large extent (if not entirely) replaced by active learning activities, class time is largely reserved for productive use of knowledge. Finally, class attendance must be mandatory. In-class instruction is geared towards promoting conceptual understanding, which is a crucially important aspect of learning. Therefore, a “flipped” classroom that adopts an optional attendance policy is not genuinely flipped. It resembles an online class more than a flipped class, since the instruction is already offered online and hence a student can afford not to attend class. In other words, pre-class study and in-class activities should be integrated. They complement each other and are

integral parts of learning as a whole.

Over the years, researchers have proposed several benefits of flipped instruction. Since many flipped classrooms use online videos to stage instruction before class, students can learn the material anytime, anywhere, and at their own pace (McDonald & Smith, 2013), which is a desired quality that has been attributed to online instruction (Bourne, Harris & Mayadas, 2005). In addition, some studies have shown that pre-class learning reduced perceived difficulty of a course, while raising perceived instructional clarity and the overall course value (Narloch, Garbin, & Turnage, 2006; Stelzer, Gladding, Mestre, & Brookes 2009). It is conceivable that by offloading some learning material to pre-class, flipped instruction effectively prepares students to engage with more challenging content during class time. Moreover, case studies have shown that flipped instruction increased teacher-student interaction during class and that students were more actively involved in the learning process in flipped classes (Herreid & Schiller, 2013).

Despite the enthusiasm, many of the studies hitherto published on flipped instruction were conducted in a single classroom with no reference to a comparison group and the purported benefits of flipped instruction have thus largely come from cases studies, descriptive surveys, theorized conjectures, and anecdotal evidence. Empirical studies employing validated controls with reasonably large sample sizes that directly assess the effectiveness of flipped instruction are sorely needed. Our study aims to address this issue. To begin with, we will review relevant research assessing the treatment effect of flipped instruction on student exam performance. Based on the review, our research questions are proposed.

1.1 Prior research on treatment effect

Several studies have reported surprisingly large effect sizes of flipped instruction.

Deslauriers, Schelew, and Wieman (2011) compared two sections of a first-year physics course taken by 538 undergraduate engineering students at University of British Columbia (UBC). Two experienced professors, one for each section, taught with traditional lectures in a similar manner. Flipped instruction was implemented only in the twelfth week of the semester, when a postdoctoral researcher with limited teaching experience replaced one professor. Students' pretest measures were practically identical. A twelve-item posttest was administered and results showed that the average scores were 41% ($SD = 13\%$, $N = 171$) in the control section and 74% ($SD = 13\%$, $N = 211$) in treatment, which gives a staggering effect size of about 2.5 standard deviations. Moreover, during the twelfth week, class attendance increased by 20% and in-class student engagement nearly doubled in the experimental section, whereas both measures remained unchanged in the control. These results are interesting in that they imply that even "novice" flipped instructors could outperform their more experienced traditional counterparts in raising student performance, class attendance and engagement.

Such encouraging results, however, should be viewed with caution. To start with, the authors themselves cautioned that the immediate end-of-treatment posttest primarily reflected the result of learning achieved from pre-class study and the class itself. Other studies with smaller effect sizes often measured student performance with end-of-term final exams that reflected all the learning done inside and outside of the class. Since learning is a multi-faceted process, the impact of a specific learning channel is likely to be diluted when more opportunities are open for students to acquire knowledge. Another contributing factor could be a sense of novelty induced

by the presence of a young and energetic instructor with a distinctly different teaching style. If the flipped pedagogy had been assessed in a more authentic setting (i.e., same instructor throughout, class flipped over an entire semester, and performance measured using end-of-term high-stakes final exams), it is unclear if the same positive results would have been achieved.

Partial evidence for the above conjecture was in fact provided by research from the same authors. In another study (Deslauriers & Wieman, 2011), flipped instruction was implemented for the entire duration of an eleven-week summer course with two cohorts of students at UBC. A superb, award-winning lecturer taught a traditional classroom in 2008. The following summer, an unspecified instructor taught the same course using the flipped pedagogy. In an ungraded mock-up exam given one week prior to the final, 62 treatment and 48 control students averaged 85% ($SD = 14\%$) and 67% ($SD = 18\%$) respectively, which corresponds to an effect size of 1.14 standard deviations, still quite large but less than half of the previous estimate.

Furthermore, two studies involving the same instructors teaching in authentic settings showed more moderate impact. Day and Foley (2006) assessed flipped instruction using two sections of the same course with 18 control and 28 treatment students. The reported statistics suggest an estimated effect size of 0.69 standard deviations with the final exam ($p = .055$). In another study, Mason, Shuman, and Cook (2013) assessed flipped instruction in a senior-level mechanical engineering course with 40 students in two consecutive cohorts (i.e., 20 students in each). Seventeen problem pairs from quizzes, midterms and finals were matched and used to assess performance. The overall effect size was 0.75 standard deviations. Although an effect size of 0.6–0.8 is still quite large (Cohen, 2013), the limited sample sizes of the two studies make

their results somewhat less convincing.

In addition, existing studies have not adequately explored possible heterogeneity of treatment effect on student performance. Mason et al. (2013) looked into the differentiated impact of flipped instruction by question type and found that their flipped classrooms were about twice as effective in improving student performance on design-based problems compared to non-design-based ones. Not enough research has been conducted examining the possibly differentiated impact of flipped treatment by student demographics, such as gender, ethnicity, and prior grades.

Finally, most studies have not carefully equated the overall instruction time and/or measured the out-of-class study time. So far, only two studies (Day & Foley, 2006; Street, Gilliland, McNeil, Royal, 2015) have explicitly controlled for instruction time by reducing the number of class meetings commensurate with the added hours of pre-class instruction. Although directly manipulating class time might encounter administrative difficulties, researchers should measure and compare out-of-class study effort whenever possible. We believe that measuring study effort in examining the benefits of flipped instruction is important. If positive treatment effects are mainly due to increase in study time (e.g., induced by the mandatory pre-class study activities), the flipped pedagogy may be less appealing. Few studies have actually compared overall study time.

In summary, although a number of empirical studies have assessed flipped instruction, with some having reported large effect sizes, compelling evidence for the benefit of flipped instruction is relatively scant. Quality empirical studies conducted in authentic settings (i.e., over

extended periods of time with high-stakes final exams) using rigorous research design (i.e., same instructors, large sample sizes, verified equivalent control groups, and overall study time measured) are still needed.

1.2 Research questions

The present study was conducted to answer the following questions:

- (1) Do flipped students spend more or less time studying outside the classroom?
- (2) Do flipped students outperform their control counterparts in exams? If so, do flipped students of diverse background benefit equally?
- (3) Do flipped students favor flipped pedagogy over traditional instruction?

2. Methodology

2.1 Course description

The current study was conducted in two sections of a first-year general chemistry course taught by the same instructor at a large public university in the western United States. This is the first quarter in a three-quarter series focusing on the fundamentals of chemistry, e.g., quantum mechanics, atomic structure and bonding, hybridization etc. Having taught the course six times in two years with the traditional lecture format, the instructor used the flipped pedagogy for the first time in the fall 2013. Both sections met three times per week on Mondays, Wednesdays and Fridays. The control class was scheduled from 1:00 to 1:50 pm, and the treatment class from 2:00 to 2:50 pm. To avoid students attending alternate sections, class attendance was mandatory and was recorded via clicker questions, which accounted for 5% of the final grade.

The control section was taught in a traditional lecture format. Although all recommended

reading assignments were posted from the beginning of the quarter, no specific reading was assigned before class. Class attendance was recorded using one iClicker question per meeting. The instructor gave lectures using presentation slides and demonstrated problem solving using a document camera. Other than the instructor occasionally pausing for questions, no significant active learning components were involved in the control class. Students completed for-credit homework assignments after class and attended discussion sessions on a voluntary basis.

Students in the flipped section were required to watch about two videos ($M = 1.96$, $SD = 0.932$) before each class meeting and complete associated assignments before class. The videos averaged 10 minutes in length ($M = 10.23$, $SD = 5.22$), most within the range of 5 to 20 minutes (range: 2–23 minutes). There were 49 videos totaling 501.27 minutes overall. The instructor created all of the videos. It took about 4 hours to produce a 10-minute video, including the time needed to prepare assignments associated with each video. To ensure compliance with pre-class preparation, four unannounced quizzes testing on pre-class material were randomly scheduled throughout the quarter. The quizzes accounted for 5% of the total grade.

In class, a typical flipped meeting was divided into three sessions. The instructor briefly reviewed pre-class material and went through each assignment for 10 to 15 minutes. The review did not repeat factual information from the videos, but instead fostered conceptual understanding by drawing connections among concepts. Students were encouraged to ask questions during this time. After the review, the instructor spent another 10 to 15 minutes with two relatively simple problems. Students worked on the problems individually and submitted their answers via clickers. For the rest of the class time (about 25 minutes), students worked on two to three increasingly

difficult worksheet problems. Solving worksheet problems required integration of multiple concepts through multiple steps. The instructor projected the problems onto a big screen and asked students to form ad hoc groups of two or three to solve problems collectively. Meanwhile, the instructor and teaching assistants walked around the classroom and offered help whenever needed. Students submitted answers via clickers and the results were dynamically displayed to the instructor. When the class faltered, the instructor provided more hints, showed polling results to foster discussion, or paused the activities to address common mistakes.

2.2 Participants

In total, 781 students were originally enrolled in the two sections, in which 54 were under the age of 18 and were excluded from the study based on our IRB stipulations. Among the eligible students, 343 (93.46%) control and 334 (92.78%) treatment students gave their written consent for participation. Student demographic information was directly collected from the University's Registrar. Table 1 shows a detailed breakdown by experimental conditions. In the combined sample, about 60% were female. Students came from 38 different majors and 13 ethnic groups. For convenience, majors were regrouped into 61.00% Biology/Chemistry, 10.49% other STEM (i.e., including all STEM majors except for biology and chemistry related ones), 4.14% Non-STEM, and 24.37% Undeclared. Similarly, ethnicity were regrouped into 11.52% White, 25.55% Black/Latino, 34.71% South Asia including Vietnamese, Thai, and Filipino, 24.67% East Asia including Korean, Chinese, and Japanese, and 3.55% Unstated. Freshmen constituted 86.12% of the students with 10.04% sophomores, 2.22% juniors, and 1.62% seniors. The average SAT math score was 591.54 ($SD = 77.32$).

2.3 Measures

2.3.1 Examinations. Two non-cumulative midterms and one cumulative final exam were delivered over the course of ten weeks. All exams were similar in form, comprised primarily of short- and long-answers. Raw scores were converted into percentages for the ease of comparison and interpretation. The treatment section took the midterms after the control section. To avoid cheating, different forms of the midterms were used with isomorphic questions. An identical final exam was given at the same time to the two sections in different rooms. Each midterm made up 23% of the total grade and final exam 36% (homework constituted the remaining 8%).

2.3.2 Main surveys. Two main surveys (i.e., a pre-survey and a post-survey) were delivered in the beginning and end of the quarter. The post-survey used in the treatment section is shown in the Appendix. Survey responses were kept away from the instructor and the teaching assistants and were not processed until after the quarter. To encourage participation, 0.5% extra credit was rewarded for completing each main survey. The survey response rates were 88.29% from the control and 86.33% from the treatment group for the pre-survey, and 82.80% and 82.92% for the post-survey.

All survey items were framed on 6-point scales from one (most negative) to six (most positive). An identical pre-survey was delivered to both sections, asking about students' perceived effectiveness of different instructional avenues (i.e., textbooks, lectures, discussion sessions, homework, and learning from peers), and about their general motivation towards this course. Our motivation items were adapted from the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993). Only the items on task

value (including interest and utility) and self-efficacy from the MSLQ were used in our study, as suggested by the expectancy-value theory (Wigfield & Eccles, 2000). Each sub-construct, i.e., interest, utility, self-efficacy, was measured by two items, with the reliability of each sub-construct as measured by Cronbach's alpha above .70 in the pre-survey and above .80 in the post-survey. Finally, the six items were averaged to produce a composite score as a higher order construct to measure general motivation with an overall alpha of .79 and .85 in the pre- and post-survey.

The pre-survey also had two questions asking about the estimated number of hours per week students spent studying before and after class for a typical mathematics or science course. In the post-survey, for the flipped condition, two items measured perceived clarity of instructional videos with a reliability of .86, and three items measured perceived instructional quality in class with a Cronbach's alpha of .88. Another two items measured students' preference of flipped instruction over traditional lectures with an alpha of .90. Two optional open-ended questions asked for students' opinions about the strengths and weaknesses of our implementation of the flipped pedagogy and their suggestions for improvement.

2.3.3 Mini-surveys. Ten identical mini-surveys, one for each week, measured students' study time outside the classroom. The mini-survey is included in the Appendix. All mini-surveys were delivered on Mondays. Each mini-survey had only two questions, asking students to give numeric estimates about the number of hours they had spent studying before and after class in the preceding week. To encourage participation, 0.1% extra credit was rewarded for completing each mini-survey. As a result, students could earn up to two extra percentage points for completing all

surveys (i.e., two main surveys and ten mini-surveys). Despite the short length of the mini-surveys, the average response rates were considerably lower for the control ($M = 68.63\%$, $SD = 7.06\%$) and treatment ($M = 66.77\%$, $SD = 8.38\%$) sections. It should be cautioned that replying on students to self-report study hours might result in over-estimation, since students are likely to exaggerate their study effort. Due to concerns with the reliability of study time measures, we did not use them in OLS regression analysis. Instead, study times were used only for between-group comparison purposes. Although the estimated study hours might be biased, the comparison could still be valid as long as students from the two sections have similar propensity for distorting their estimates.

3. Results

3.1 Group equivalence

Table 1 shows the descriptive statistics by section on survey responses, exam performance, and demographics. Reasonable group equivalence was found by all measures except for major and ethnicity. Chi-squared test showed statistically significant ($p = .021$) differences in major. It can be argued, however, that the size of the difference is not practically important. The flipped section had 6.51% less STEM majors, 3.09% more Chemistry/Biology majors, 2.47% more Non-STEM majors, and 0.94% more Undeclared majors. Such differences are reasonably small and should not raise serious concerns. Similarly, differences in ethnicity between groups are even smaller and the overall chi-squared test is only marginally significant ($p = .069$). These small p values are most likely due to our relatively large sample size (i.e., 677 in total), which tends to produce small standard errors that accentuate differences.

3.2 Out-of-class study time

(1) Did flipped students spend more or less time studying outside the classroom?

Trajectories of study time are shown in Figure 1 with descriptive and test statistics shown in Table 2. Effect sizes (ES) associated with two-sample t -tests are computed using Cohen's d . Two local maximums at weeks 4 and 8 correspond to the first and second midterms, which explains the dips in average study time immediately afterwards. Flipped students on average spent more time before class ($ES = 0.146, p = .056$), especially in the first nine weeks ($ES = 0.158, p = .046$) and less so during the tenth week ($ES = 0.025, p = .774$). In contrast, flipped students tended to spend less time after class ($ES = -0.152, p = .047$), particularly during the tenth week ($ES = -0.168, p = .046$). Taken together, the overall out-of-class study time was roughly the same across weeks ($ES = -0.026, p = .742$). To note, the self-reported study times were highly skewed to the right. Therefore, we have conducted sensitivity analysis by removing or truncating extreme values, which resulted in smaller standard errors leading to the same conclusions with more distinct patterns.

3.3 Exam performance

(2) Did flipped students outperform their control counterparts in exams? If so, did flipped students of diverse background benefit equally?

As shown in Table 1, a simple two-sample t -test suggests that flipped instruction had a small but statistically non-significant effect on the final exam ($ES = 0.116, p = .132$). Although the combined sample size is large, due to the small mean difference in the final exam and the magnitude of the standard deviations, the statistical power computed *a posteriori* is only about

33%. In other words, the two-sample t -test does not have enough statistical power to detect the small treatment effect as observed. To investigate further, OLS regression models were used to include additional covariates, leading to smaller residual errors and thus more power to detect small differences. Another reason for using OLS regression is to account for minor imbalances as shown previously in group-equivalence check. The results are shown in Table 3. To note, all regression coefficients presented in this study are standardized beta coefficients (β) that can be readily interpreted as effect sizes (ES). The first two models used the final exam scores as the dependent variable. Model 3.1 is the main effect model, which includes all statistically significant covariates without interaction terms. It shows a small but statistically significant treatment effect ($\beta = 0.189, p = .010$). From Model 3.1, potential interaction terms were studiously explored and the only significant interaction found is between SAT math and gender ($\beta = 0.197, p = .015$), which is shown in Model 3.2 with larger adjusted R square and smaller AIC index. Post-regression diagnostics on Model 3.2 did not show noticeable violations of OLS assumptions; nor were any extremely influential cases identified. Therefore, Model 3.2 is accepted as our final model.

Closer inspection of Table 1 indicates that the treatment effect on exam performance was the strongest with the first midterm, disappeared by the second, and rebounded slightly with the final exam. It seems that the net benefit of flipped instruction diminished over time. As a result, Models 3.3–3.6 were fitted to investigate the mediating effect of the first midterm. Conceptually, the treatment effect on performance in Model 3.2 can be regarded as the total effect. We want to examine if the first midterm can be treated as a mediator. If so, we should see significant impact

of treatment effect on the first midterm, but non-significant impact of treatment effect on the final exam with the first midterm included as a covariate.

Model 3.3 resembles Model 3.2 in all aspects except that the dependent variable was the first midterm scores. Model 3.4 improves upon Model 3.3 with significant interaction terms added and non-significant covariates removed. Models 3.5 and 3.6 used final exam scores as the dependent variable with the first midterm included as a covariate. Model 3.5 shows that (a) doing well in the first midterm strongly correlated with the outcome in the final ($\beta = 0.626, p < .001$), (b) the direct effect of flipped instruction on final exam scores while controlling for the first midterm was negligibly small ($\beta = 0.022, p = .696$); and (c) the total effect of flipped instruction ($\beta = 0.192$) as shown in Model 3.2 was primarily due to the mediating effect of the first midterm (i.e., $0.157 = 0.248 \times 0.632$).

3.4 Perception and attitude

(3) Did flipped students favor the flipped pedagogy over traditional instruction?

All participating students in the treatment condition were asked to rate their attitudes towards flipped instruction in the post-survey. The overall flipped students' attitude was distinctly lukewarm. Students' ratings on their preference of flipped instruction over traditional lectures averaged 3.631 ($SD = 1.538$) on a 6-point scale, which is not significantly different from the neutral position of 3.50 ($p = .156$). A histogram of the preference measure reveals a broadly bell-shaped distribution with the global maximum around the mean and two local maximums peaking at the two extremes. This result indicates that opinions towards the flipped pedagogy were somewhat bipolar, as there were about as many students who strongly favored flipped

instruction as those who fiercely opposed it (i.e., 30 students or 10.83% rated six and 34 students or 12.27% rated one on the 6-point scale).

To gain insight as to whether prior motivation and student demographics associate with perceived class quality and instructional clarity, we applied OLS regression and the results are shown in Table 4. With all students included, Model 4.1 shows that flipped instruction did not raise perceived in-class quality. In contrast, students with stronger prior general motivation ($\beta = 0.312, p < .001$) and higher SAT math scores ($\beta = 0.217, p < .001$) tended to perceive in-class activities to be of higher quality. Moreover, non-Chemistry or non-Biology majors generally gave lower ratings. Likewise, flipped students' perception of video clarity demonstrated similar patterns of association, as students with higher SAT math scores and motivation rated instructional clarity more positively. Although major is not a statistically significant predictor, it is kept in the model for the sake of comparison. To note, the size of the coefficients for major in Model 4.2 is about as large as that in Model 4.1. The halved sample size increased standard errors, which might push the coefficients out of significance.

In the flipped section, 82.92% of the students responded to the post-survey, and about half of them opted to express their likes and dislikes about the flipped pedagogy. The comments have offered valuable insights. Not surprisingly, student comments echoed their bipolar ratings. Some expressed unqualified approval, while others showed bitter resentment. Specifically, positive comments have confirmed some of the proposed benefits of flipped instruction, which include (a) the flexibility of “*watching videos at my own pace, whenever, wherever, and however many times I would like*”; (b) improved preparedness for class, as it is “*a means to make lecture*

portion much easier to understand” and “helping me feel more confident for knowing what the lecture is about, so you can just accumulate knowledge from that point”; and (c) capacity to accommodate more elaboration, application, and teacher-student interaction, since “by getting the dry stuff out of the way, the inverted method enables more teacher-student time during the classes,” and “we were able to do more practice problems in class and that really cleared things up for me.”

On the other hand, those who opposed flipped instruction voiced strong criticism. Some students were accustomed to the old ways and resented changes. *“When I attend a lecture, I wanted a professor to actually teach me the content. If I wanted to learn chemistry online then I could YouTube it myself rather than someone telling me, but since I'm paying for the course I feel that it would be more suitable for the professor to lecture during class.”* Some students did not understand that the pre-class and in-class instructions are inseparable and could not see the value of instructor-guided problem solving in class. *“I did not like it at all, the inverted way made me feel like it was a waste of time to even go to class. I can do practice problems on my own time.”* The bulk of the criticism was leveled at video and class related issues; *“I disliked the voice quality of the videos. Some online videos had a very muffled voice from an older microphone”; “Sometimes it felt like some things were rushed or that we spent too much time on one thing”; and “The practice problems in class seem to be easier compared to midterms' questions. I would have appreciated more exam style practice problems during lecture with the same level of difficulty of an exam.”*

4. Discussion

4.1 Out-of-class study time

Several studies have reported without sufficient proof that students claimed to have spent no more, or even less time, studying in flipped classrooms (Foertsch, Moses, Strikwerda, & Litzkow, 2002; Mason et al., 2013; Narloch et al., 2006), except for one study reporting the opposite (Papadopoulos & Roman, 2010). Our data verified the claim that flipped instruction does not appreciably increase the overall workload of the students. This suggests that to assess the effectiveness of flipped instruction, it may not be necessary to adjust additional pre-class study time by reducing the number of class meetings (Street et al., 2015). The result also implies that any treatment effect on exam performance should be attributed to factors other than mere increase in study effort. One possibility is that flipped students might benefit from spaced learning, since they allocated their study time more evenly than regular students did. Given the same amount of time, spaced learning leads to better retention of information, a phenomenon known as the spacing effect (Donovan & Radosevich, 1999).

4.2 Exam performance

Our regression models suggest a small but statistically significant ($\beta = 0.192, p = .008$) effect of flipped instruction on student final exam performance. Compared to the effect sizes reported from other studies (Deslauriers, et al., 2011; Deslauriers & Wieman, 2011; Moravec et al., 2010), our result is appreciably smaller. We believe, however, an effect size of about 0.2–0.4 is much more likely in practice than some large effect sizes previously reported. As discussed earlier, studies conducted in relatively short periods of time using immediate end-of-term, low-stakes tests could arguably be more likely to produce more favorable results for several

reasons. First, the novelty induced by a distinctly different instructional technique could temporarily intrigue and motivate students. Second, in shorter time periods, fewer things could possibly go amiss and hence make it more likely for a complex instructional technique such as flipped instruction to work, which entails making multiple decisions on pre-class and in-class components. In a flipped classroom, for example, an instructor need to consider the number and length of videos, accompanying practice questions, pre-class quizzes, percentage of lectures retained in class, the number and kinds of in-class active learning activities to adopt, and different ways to conduct them. The more decisions to make, the more it is likely that some steps can go wrong. Third, using immediate, low-stakes tests makes it more difficult for other compensating mechanisms to work, hence overestimating the impact of flipped instruction. The final exam often accounts for the most of a student's grade. Therefore, students in both treatment and control conditions will do extra activities in their own time to prepare for it, which will somewhat drown out any positive impact brought about by flipped instruction. The residual impact, after alternative overcompensating learning mechanisms are allowed to take effect, is therefore a more meaningful, practical measure of the overall effect of flipped instruction. As a result, we recommend using long-term, high-stakes, tests and hence regard results from cumulative final exams more highly than those from midterms.

In fact, our study has already demonstrated the possibility of diminishing treatment effect over longer time period: The effect size of flipped instruction from two-sample *t*-tests with the first midterm was about twice as large as that with the final exam. In addition, our regression results regarding the potential mediating effect of the first midterm on final exam performance

indicate that the treatment effect on the final exam was largely attributed to the lingering effect of flipped instruction on the first midterm. In other words, flipped instruction contributed little to accruing benefits after the first midterm, which is partly supported by the practically null effect ($ES = -0.050$, $p = .515$ with two-sample t -test) by the second non-cumulative midterm.

Moreover, studies similar in design to ours (i.e., assessing flipped instruction using end-of-quarter or semester finals with large sample sizes) have reported comparable effect sizes (Street et al., 2015, Wong, Ip, Lopes, & Rajagopalan, 2014). On the other hand, it should also be duly noted that this was the first time our instructor had implemented flipped pedagogy. Since the instructor had spent most of the preceding summer (totaling about 200 hours) developing instructional videos, there was less time devoted to preparing materials such as pre-class assignments and quizzes that would help ensure compliance. It is possible that the lack of enough for-credit quizzes as incentives was accountable for non-compliance, which in turn lessened treatment effect in our case.

4.3 Perceptions and attitudes

Our results have shown that treatment students had mixed, polarized feelings about flipped instruction. Moreover, Table 4 implies that generally more motivated and academically well-prepared students might be more receptive to flipped instruction, as they tended to perceive the class to be of higher quality and the instruction of greater clarity. By closely examining students' comments, we began to understand some weaknesses of our implementation of the flipped pedagogy, which is helpful for understanding previous results.

Specifically, the essence of flipped instruction is to move certain instructional material

outside the classroom to free up class time for problem-solving and teacher-student interaction.

Its success, therefore, critically hinges on the success and effectiveness of pre-class study and in-class active learning activities. In our case, non-compliance with pre-class study was a serious issue. Three causes are identified. The first is habitual resistance. Many students commented that they were completely new to flipped instruction and were not fond of learning before class.

Some customarily associated learning with instruction and firmly believed that “*being explicitly taught step-by-step is the most efficient way to learn*”. Many could not see the critical importance of pre-class study and regarded it as “*extra work*” they were forced to do rather than a mere shift in workload. Second, time management skill and self-discipline are needed. One highly desired appeal of online videos is immortalized in the reprise, *anytime, anywhere, at any pace*. The irony is that having videos readily accessible any time online induces procrastination. In traditional classrooms, class time was designated for learning. In flipped classrooms, however, students had to decide when to study on their own and some simply lacked the self-discipline to do so in a timely manner. When they studied online, some were easily distracted and lured away by other websites. Occasional lapses were a factor particularly during the second half of the quarter. For each class meeting, the minimum study time was about 20 minutes to simply play the videos. As time went on, the material became more challenging and pre-class study time multiplied.

Meanwhile, other courses also became increasingly demanding towards the end, making unintended non-compliance more likely to happen. Once students missed watching the videos, attending class would not be as helpful. Some were clueless during group discussion and frustration led them to simply sit out the time in class rather than using the time to catch up. For

whatever reasons, when students failed to adequately prepare before class and hence fared poorly in class, complaints would ensue and radiate in multiple directions (e.g., video examples could be more difficult; instructor should review video contents in class; instructor went through solutions too fast). By implication, non-compliance seemed to disproportionately affect students with low motivation, poor self-discipline, and weak time-management and academic skills.

The negative impact of non-compliance from some students rippled to affect others as well. While reading the comments, we identified an interesting association: If a student reported having enjoyed the videos and benefited from studying before class, the student would often lament the lack of enough challenging practice problems, complaining that the class time was not productively spent. In contrast, a major complaint voiced by the control students was that the instructor was too fast, too rushed, and covered too much content. Therefore, the question is why the instructor would not simply give more challenging practice problems to the treatment students. As one flipped student had observed, *“During office hours, Dr. B would take aside 10-20 students and teach them. Students were quiet and attentive while she answered their questions. Dr. B's explanations were concise and she was able to answer the follow-up questions in-depth. During lectures, Dr. B was forced to water down her explanations so that the least informed of the four hundred students could probably understand. The lectures usually went at the speed of the slowest student.”*

We believe that non-compliance is not only the root cause for various complaints, but also sheds light on the overall small treatment effect, absence of marked interaction, and diminishing treatment effect. Non-compliance affected both under-prepared and well-prepared

students, as it made class activities less useful for the under-prepared while limiting the amount and difficulty of the practice problems that could have been solved in class. In other words, non-compliance hinders flipped instruction from reaching its full potential. The overall treatment effect was hence small.

With regard to interaction, we believe that an interaction effect occurs when the treatment conditions clearly agree with the characteristics of a specific subgroup. Others with characters departing from this niche group in varying degrees consequently enjoy the benefits to lesser extents. In our study, since neither well-prepared nor under-prepared students had enjoyed the full benefits of flipped instruction, no obvious niche group could be identified and hence no marked interaction was detected.

We also believe that non-compliance may have been amplified by the class composition, which included 86% freshmen. Due to the time spent on developing videos, the instructor did not create enough for-credit quizzes to ensure compliance. It is conceivable, when left to their own devices, freshmen might be particularly vulnerable, since they were likely to have poor self-discipline, weak time-management and learning skills, compared to their more senior counterparts. Towards the end of the quarter, unintended non-compliance might become increasingly common, which in turn would translate into diminishing treatment effects in later weeks. Moreover, student attitudes did not remain static, but tended to grow towards the extremes over time, which eventually manifested as polarized feelings in the post-survey.

The conjecture that freshmen might be less receptive to flipped instruction has also been suggested elsewhere. Mason et al. (2013) assessed flipped instruction in an upper-level

engineering course taken exclusively by seniors. By the fourth week of the quarter, students already agreed that flipped instruction was “a better use of class time”. Most surprisingly, however, the end-of-quarter survey indicated that none of the 20 seniors believed that the flipped pedagogy would work in first-year courses, since the freshmen “lacked the academic maturity needed to succeed in this setting”. As a result, the authors conceded that flipped instruction “may be difficult for students who have not developed strong study skills.”

5. Limitations

While response rates to the main surveys were reasonably high, the weekly mini-surveys had much lower rates, averaging 67.7% ($SD = 7.7\%$). The low response rates could result from the absence of reminder emails. For main surveys, response rates always jumped at the prompt of each reminder email. Since mini-surveys were delivered weekly, sending reminder emails was considered overly intrusive. If frequent responders to mini-surveys were more likely to possess greater self-discipline and time management skills essential for adequate pre-class study, our measured study times would be biased.

In addition, despite our efforts to ensure students that survey responses would not be analyzed until after the quarter, students could still be motivated to exaggerate their study effort, a cause for upwardly biased study times. Although students with different characteristics might have varying propensities for over-estimation, between-group comparison of overly estimated study times might still be valid as long as students from the two sections have comparable characteristics and hence similar propensities for over-estimation. Future studies should consider alternative ways to measure out-of-class study time more accurately.

External validity is also a matter of concern. Student demographics in our study were over 60% Asian, about 12% Caucasian, and less than 2% Black or African-American in the combined sample. This composition is clearly different from many other institutions in the US and across the globe. It is unclear how much the ethnicity mix would impact the generalization of our results and findings.

6. Conclusions and Implications

Flipped instruction did not increase students' overall study time; it only caused a shift in student workload. By implication, any impacts of flipped instruction should be attributed to factors other than mere increase in study effort. Moreover, to assess flipped instruction, it might be unnecessary to adjust additional pre-class study time by reducing the number of class meetings. By allocating study time more evenly, treatment students might in theory benefit from spaced learning.

We believe measuring student performance using long-term, high-stakes exams gives more practically meaningful results. With flipped instruction implemented for the first time, our OLS models showed a small, but statistically significant, treatment effect ($ES = 0.192, p = .008$) with the final exam. No marked interaction was identified, indicating flipped instruction benefited all students equally. The overall treatment effect was more pronounced in the beginning, but diminished over time.

Flipped instruction did not increase student motivation and perceived overall class quality. Treatment students' preference of flipped instruction over traditional lectures was lukewarm with about one fifth of the students displaying polarized feelings. Prior motivation and SAT math

scores were positively associated with perceived class quality and instructional clarity, which suggests that highly motivated and academically well-prepared students might be more receptive to flipped instruction.

Positive student comments confirmed some proposed benefits of flipped instruction, including learning at one's own time and pace, better preparation for class, and more problem solving and teacher-student interaction. Pre-class study non-compliance was a serious issue in our study and was believed to be closely associated with negative student attitudes. On students' part, three possible causes of non-compliance were identified, i.e., habitual resistance, procrastination and distraction, and unintended lapses. By implication, non-compliance seemed to disproportionately affect students with poor self-discipline, low motivation, and weak time-management and academic skills. Moreover, non-compliance seemed to affect all students, as under-prepared students tended to have difficulty following the class while well-prepared students reported boredom and demanded for more challenging practice problems during the class. As a result, the overall treatment effect may have diminished and no marked interaction was detected (since no niche group had enjoyed the full benefits of flipped instruction). **The predominance of freshmen in the class may also lead to diminished treatment effect.** Since not enough for-credit quizzes were created to ensure compliance, freshmen were particularly vulnerable due to weak self-discipline, time-management, and learning skills. Unintended lapses became increasingly frequent towards the end of quarter, leading to diminished treatment effect.

In summary, our implementation of the flipped pedagogy caused a shift in student workload from post-class to pre-class study without appreciably increasing the overall amount.

Flipped instruction slightly, but uniformly, increased final exam performance for all subgroups of students. However, it did not increase student motivation and perceived overall class quality with about one fifth of the students showing polarized feelings. Non-compliance to pre-class study, lack of enough pre-class quizzes, and the predominance of freshmen in class are believed to result in the diminished treatment effect and absence of marked interaction. Future practitioners of flipped instruction should take measures to ensure pre-class study compliance, particularly in large introductory undergraduate courses, as the success and effectiveness of flipped instruction critically depends on both the pre-class and in-class components.

References

- Bergmann, J., & Sams, A. (2008). Remixing chemistry class. *Learning and Leading with Technology*, 36(4), 24-27.
- Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. In *ASEE National Conference Proceedings, Atlanta, GA*.
- Bourne, J., Harris, D., and Mayadas, F., (2005). Online engineering education: learning anywhere, anytime. *Journal of Engineering Education*, 94(1), 131-146.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Day, J. A., & Foley, J. D. (2006). Evaluating a web lecture intervention in a human-computer interaction course. *Education, IEEE Transactions on*, 49(4), 420-431.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332(6031), 862-864.
- Deslauriers, L., & Wieman, C. (2011). Learning and retention of quantum concepts with different teaching methods. *Physical Review Special Topics - Physics Education Research*, 7(1), 010101-1-6.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84(5), 795-805.
- Foertsch, J., Moses, G., Strikwerda, J., & Litzkow, M. (2002). Reversing the lecture/homework paradigm using eTEACH® web-based streaming video software. *Journal of Engineering Education*, 91(3), 267-274.
- Herreid, C. F., & Schiller, N. A. (2013). Case studies and the flipped classroom. *Journal of College Science Teaching*, 42(5), 62-66.

- Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, 31(1), 30-43.
- Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course. *Education, IEEE Transactions on*, 56(4), 430-435.
- McDonald, K., & Smith, C. M. (2013). The flipped classroom for professional development: part I. Benefits and strategies. *The Journal of Continuing Education in Nursing*, 44(10), 437.
- Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before lecture: a strategy that improves learning outcomes in a large introductory biology class. *CBE-Life Sciences Education*, 9(4), 473-481.
- Narloch, R., Garbin, C. P., and Turnage, K. D. (2006). Benefits of prelecture quizzes. *Teaching Psychology*, 33(2), 109-112.
- Papadopoulos, C., & Roman, A. S. (2010). Implementing an inverted classroom model in engineering statics: Initial results. In *117th American Society for Engineering Education. American Society for Engineering Education*. Louisville, Kentucky, USA.
- Pintrich, P. R., Smith, D. A., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement*, 53(3), 801-813.
- Stelzer, T., Gladding, G., Mestre, J. P., & Brookes, D. T. (2009). Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content. *American Journal of Physics*, 77(2), 184-190.

- Street, S. E., Gilliland, K. O., McNeil, C., & Royal, K. (2015). The flipped classroom improved medical student performance and satisfaction in a pre-clinical physiology course. *Medical Science Educator*, 25(1), 35-43.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25(1), 68-81.
- Wong, T. H., Ip, E. J., Lopes, I., & Rajagopalan, V. (2014). Pharmacy students' performance and perceptions in a flipped teaching pilot on cardiac arrhythmias. *American Journal of Pharmaceutical Education*, 78(10), 1-6.

Table 1

Descriptive Statistics of Survey Responses, Exam Performance, and Demographics by Section

Measure	Control (<i>N</i> = 343)	Treatment (<i>N</i> = 334)	<i>t</i> (<i>p</i>) or χ^2 (<i>p</i>)	Cohen's <i>d</i>
	<i>M</i> (<i>SD</i>) or Percentage	<i>M</i> (<i>SD</i>) or Percent		
Motivation (pre-survey)	4.63 (0.75)	4.68 (0.83)	0.82 (0.414)	0.064
Before-class Effort (pre-survey)	4.93 (6.01)	4.91 (5.42)	-0.03 (0.977)	-0.003
After-class Effort (pre-survey)	7.27 (7.85)	7.75 (9.55)	0.66 (0.507)	0.055
Textbook (pre-survey)	4.98 (2.82)	5.08 (2.82)	0.43 (0.666)	0.035
Lecture (pre-survey)	4.49 (2.60)	4.71 (2.52)	1.08 (0.281)	0.086
Discussion Session (pre-survey)	4.46 (2.74)	4.59 (2.59)	0.62 (0.532)	0.049
Homework (pre-survey)	4.70 (2.87)	4.79 (2.72)	0.40 (0.692)	0.032
Peers (pre-survey)	4.90 (3.04)	5.24 (2.81)	1.37 (0.170)	0.113
SAT Math	595.23 (75.76)	587.79 (78.81)	-1.21 (0.226)	-0.096
Freshman	86.30%	85.92%	3.91 (0.272)	
Sophomore	10.79%	9.28%		
Junior	1.17%	3.29%		
Senior	1.75%	1.50%		
Chemistry/Biology	59.48%	62.57%	9.69 (0.021)	
STEM	13.70%	7.19%		
Non-STEM	2.92%	5.39%		
Undeclared	23.91%	24.85%		
Female	60.06%	61.26%	0.10 (0.749)	
Black/Latino	23.91%	27.25%	8.71 (0.069)	
East Asia	25.36%	23.95%		
South Asia	32.65%	36.83%		
Unstated	3.21%	3.89%		
White	14.87%	8.08%		
Midterm1	49.90 (21.48)	54.38 (22.80)	2.63 (0.009)	0.202
Midterm2 (non-cumulative)	64.57 (18.79)	63.66 (17.49)	-0.65 (0.515)	-0.050
Final Exam (cumulative)	42.5 (12.51)	43.9 (11.59)	1.51 (0.132)	0.116
Motivation (post)	4.13 (1.14)	4.24 (1.10)	1.22 (0.222)	0.098
Class Quality (post)	4.03 (1.10)	4.05 (1.12)	0.27 (0.785)	0.018
Prefer Flipped Instruction (post)		3.63 (1.54)		

Table 2

Self-reported Out-of-class Study Time in Hours by Section

	Week	Control Mean (SD)	Treatment Mean (SD)	<i>t</i> -statistic (<i>p</i>)	Cohen's <i>d</i>
Before-class	Weeks 1-10	3.71 (3.11)	4.14 (2.76)	1.92 (0.056)	0.146
	Weeks 1-9	3.55 (2.87)	3.98 (2.56)	2.00 (0.046)	0.158
	Week 10	5.12 (5.29)	5.24 (5.17)	0.29 (0.774)	0.023
After-class	Weeks 1-10	6.28 (4.16)	5.69 (3.58)	-1.99 (0.047)	-0.152
	Weeks 1-9	6.00 (4.01)	5.51 (3.47)	-1.65 (0.100)	-0.131
	Week 10	8.07 (6.49)	7.01 (6.09)	-2.00 (0.046)	-0.168
Out-of-class	Weeks 1-10	9.99 (6.64)	9.83 (5.85)	-0.33 (0.742)	-0.026
	Weeks 1-9	9.54 (6.29)	9.48 (5.43)	-0.12 (0.901)	-0.010
	Week 10	13.19 (10.88)	12.26 (10.18)	-1.05 (0.294)	-0.088

Note. Averaged study times were used for producing the estimates. Removing or truncating extreme values generated rather similar results leading to the same conclusions.

Table 3

Effect of Flipped Instruction on Exam Performance with OLS Models

	Final Exam		First Midterm		Final Exam	
	Model 3.1	Model 3.2	Model 3.3	Model 3.4	Model 3.5	Model 3.6
(Intercept)	0.302*	0.333**	0.333**	0.388***	0.124	0.043
	(0.123)	(0.124)	(0.125)	(0.114)	(0.096)	(0.060)
Treatment	0.189**	0.192**	0.273***	0.248***	0.022	0.014
	(0.072)	(0.072)	(0.073)	(0.068)	(0.056)	(0.056)
First Midterm					0.626***	0.632***
					(0.033)	(0.031)
Motivation (pre)	0.086*	0.087*	0.046		0.058*	0.061*
	(0.036)	(0.036)	(0.036)		(0.028)	(0.028)
SAT Math	0.477***	0.36***	0.340***	0.048	0.148**	0.128**
	(0.044)	(0.065)	(0.066)	(0.130)	(0.051)	(0.048)
Female	0.031	0.026	0.045		-0.001	-0.011
	(0.08)	(0.079)	(0.08)		(0.061)	(0.060)
SATmath×Female		0.197*	0.115		0.124*	0.122*
		(0.081)	(0.082)		(0.062)	(0.062)
STEM	-0.035	0.038	-0.123	-0.154	0.115	0.064
	(0.132)	(0.134)	(0.136)	(0.121)	(0.104)	(0.100)
Non-STEM	-0.625**	-0.552*	-0.487*	-0.508*	-0.248	-0.316*
	(0.231)	(0.232)	(0.235)	(0.223)	(0.180)	(0.157)
Undeclared	-0.312***	-0.302***	-0.336***	-0.297***	-0.094	-0.108
	(0.089)	(0.088)	(0.089)	(0.085)	(0.069)	(0.069)
Sophomore	0.101	0.098	0.292*	0.283*	-0.085	
	(0.118)	(0.118)	(0.119)	(0.117)	(0.092)	
Junior	-0.113	-0.176	-0.140	-0.067	-0.089	
	(0.312)	(0.312)	(0.316)	(0.312)	(0.241)	
Senior	0.332	0.290	0.926*	0.628*	-0.290	
	(0.387)	(0.386)	(0.391)	(0.312)	(0.300)	
Black/Latino	-0.488***	-0.502***	-0.657***	-0.680***	-0.093	
	(0.131)	(0.131)	(0.133)	(0.134)	(0.104)	
South Asia	-0.140	-0.161	-0.176	-0.243*	-0.051	
	(0.125)	(0.124)	(0.126)	(0.122)	(0.096)	
East Asia	-0.494***	-0.493***	-0.487***	-0.542***	-0.189 ⁺	
	(0.131)	(0.131)	(0.132)	(0.134)	(0.102)	
Unstated	-0.260	-0.245	-0.484*	-0.621**	0.057	
	(0.229)	(0.228)	(0.232)	(0.218)	(0.177)	
SATmath×Black/Latino				0.383*		
				(0.152)		
SATmath×South Asia				0.526***		
				(0.146)		
SATmath×East Asia				0.301*		
				(0.147)		
SATmath×Unstated				0.260		
				(0.201)		
Cases	555	555	556	635	555	555
Adj.R-squared	0.305	0.311	0.3	0.3	0.588	0.588
AIC	1391.35	1387.22	1404.74	1608.96	1103.18	1096.07

Note. All estimates are standardized beta coefficients. Standard errors are in parentheses.

⁺ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Table 4

Perceived In-class Quality for All Students and Video Clarity for Flipped Students

	Model 4.1 (In-class Quality)	Model 4.2 (Video Clarity)
(Intercept)	4.117*** (0.08)	4.276*** (0.099)
Treatment	-0.018 (0.101)	
SAT Math	0.217*** (0.055)	0.231* (0.089)
Motivation (pre-survey)	0.312*** (0.052)	0.175* (0.084)
STEM	-0.348* (0.177)	-0.496 (0.366)
Non-STEM	-0.518 ⁺ (0.284)	-0.355 (0.395)
Undeclared	0.039 (0.125)	0.070 (0.208)
Cases	465	226

Note. All estimates are standardized beta coefficients. Standard errors are in parentheses.

⁺ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Figure 1

Appendix

Post-survey (for treatment section)

1. Based on your learning experience in this quarter, please rate the following items.

<i>Please rate your agreement with the following statements</i>	Strongly Disagree				Strongly	
I am very interested in the content area of this course.	1	2	3	4	5	6
Beyond this quarter, this course will still be very useful to me.	1	2	3	4	5	6
I am very confident that I will do well in this course.	1	2	3	4	5	6
I find studying the course material enjoyable.	1	2	3	4	5	6
I will need the contents from this course in subsequent courses	1	2	3	4	5	6
Given my current situation, I am confident of getting a good grade.	1	2	3	4	5	6
Professor's online videos were crystal clear to me.	1	2	3	4	5	6
Professor's in-class instruction was crystal clear to me.	1	2	3	4	5	6
I prefer this inverted class format to a traditional "lecture" format.	1	2	3	4	5	6
I would prefer to take more science classes using this type of class format.	1	2	3	4	5	6

<i>Please rate the overall quality of the following items</i>	Poor				Excellent	
online component of the instruction	1	2	3	4	5	6
in-class component of the instruction	1	2	3	4	5	6
this course as a whole	1	2	3	4	5	6

2. Open-ended Questions (optional)

2.1. How do you like or dislike about the "inverted" method of teaching this course?

2.2. What is your major complaint about this course and how do you recommend us to improve it?

Mini-survey (for both sections)

Based on your learning experience in the preceding week, please answer the following questions.

1. Before class, I spent about _____ hours in total, learning course materials in advance.
2. After class, I spent about _____ hours in total, reviewing course materials and doing homework.