International Journal of Teaching and Learning in Higher Education 2017, Volume 29, Number 3, 421-435
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Can Flipping the Classroom Work? Evidence From Undergraduate Chemistry

Timothy Casasola, Tutrang Nguyen, and Mark Warschauer
University of California, Irvine

Katerina Schenke
The University of California, Los Angeles

Our study describes student outcomes from an undergraduate chemistry course that implemented a flipped format: a pedagogical model that consists of students watching recorded video lectures outside of the classroom and engaging in problem solving activities during class. We investigated whether (1) interest, study skills, and attendance as measured by self report improved during the term as a result of course format (n=252) and (2) students in a flipped chemistry course earned higher grades in the subsequent chemistry course compared with students who enrolled in the non-flipped course that same term (n=295). Although we found no significant differences between students' self-reported interest and study skills at the end of the term, we found that students enrolled in the flipped course reported attending class more often than students in the non-flipped course (β = .32). We also found that after controlling for student-level covariates related to achievement (such as SAT Math scores and grade in previous chemistry course), students enrolled in the flipped chemistry course experienced, on average, a statistically significant increase of half a standard deviation (β = .55) in their grade in the subsequent chemistry course. We discuss implications for study of flipped instruction.

Undergraduate institutions are faced with a big problem: too few students major in STEM (science, technology, engineering, and math) fields, and too many who start these majors abandon them before graduating (see CSRDE, 2013). The question of student engagement in STEM fields is one that spans a variety of perspectives, ranging from issues of educational equality to motivation to pursue STEM careers. However, one issue that undergraduate institutions can directly attend to is promoting high-quality instruction in STEM courses (see Fairweather, 2009). Past research has found that lower-division STEM courses often focus too much on providing information and too little on fostering scientific discussion, analysis, and reflection (Baillie & Fitzgerald, 2010; see NAE Annual Report, 2005). Further, many questions have been raised about whether these courses are effective in encouraging students to continue to pursue their STEM major (Baillie & Fitzgerald, 2010; Kyle, 1997; McGinn & Roth, 1999; Mervis, 2010; NAE 2005).

In contrast, students in courses that use active engagement instructional approaches tend to drop out less and earn higher grades (Freeman et al., 2014). Active engagement approaches have also been found to improve students' conceptual understanding and attitudes toward the subject (Beichner et al., 2007). As a result, large universities are exploring new approaches to undergraduate education, seeking to identify active learning approaches that can keep students engaged and enrolled in STEM majors. As such, the present study investigates the effectiveness of one type of pedagogical approach—flipping the classroom—on student interest, study skills, and attendance in the course and student achievement in the subsequent course.

The Flipped Classroom

One approach that contains elements of active learning is the "flipped classroom" (Bishop & Verleger, 2013; Strayer, 2012). The flipped classroom inverts, or "flips," where concept absorption and concept engagement traditionally take place: in a flipped class, the majority of concept absorption happens outside of the classroom while the majority of concept engagement happens in the classroom (Bergmann & Sams, 2012). Watching recorded video lectures outside of the classroom is currently the most common instance of concept absorption in the flipped format. Concept engagement in class may look like working on problem-solving activities individually and/or in groups, sometimes with the help of peers, teaching assistants, or the instructor. Flipping the classroom allows class time to be used for students to ask questions, engage in problem solving, and practice the concepts that were learned outside of class, thus allowing students to better prepare for learning the material. Flipped instruction changes where students practice and engage with the material. Instead of applying difficult concepts outside of class in isolation, flipped instruction encourages students to apply concepts during class with guidance from the instructor and other peers.

There are some reasons to believe that flipped classrooms may benefit student outcomes more than non-flipped classes. For example, the flipped format has had positive effects on student problem solving skills (Khousmi, & Hadjou, 2005; Wilson, 2013). Additionally, students who take a flipped course earned higher grades on common exams (Baepler, Walker, & Driessen, 2014; Deslauriers, Schelew, & Wieman, 2011; He, Holton, Farkas, & Warschauer, 2016; Moravec, Williams, Aguilar-Roca, & O'Dowd, 2010; O'Flaherty & Phillips, 2015) and higher overall grades than students who take the same course in a traditional, non-flipped format (Baepler et al., 2014; Deslauriers et al., 2011; Kim, Kim, Khera, & Getman, 2014; Moravec et al., 2010; O'Flaherty & Phillips, 2015; Roach, 2014; Strayer, 2012). However, some studies found no statistically significant differences between student achievement in flipped and non-flipped classes (see Clark, 2015; Jensen, Kummer, & Godoy, 2015; Morin, Kecskemety, Harper, & Clingan, 2013; Muzyk et al., 2015, Rias-Rohani & Walters, 2014), leading to unclear conclusions about the benefits of flipped instruction on student achievement.

In addition to improvements in performance and skills, some research has shown that students have positive perceptions of flipped instruction (Stelzer, Brookes, Gladding, & Mestre, 2010; Deslauriers et al., 2011; Mason, Shuman, & Cook, 2013; Kim et al. 2014; Roach, 2014). For example, students reported that being able to watch videos on their own time aids their studying since they can pause and rewind lectures (Roach, 2014). McCallum, Schultz, Selke and Spartz (2015) conducted an exploratory study to understand students' views on academic, peer-to-peer, and student-faculty involvement in three flipped undergraduate courses. Across these three dimensions, some of the themes that emerged were lecture accessibility, engaging in-class experience, relationship building, and professor awareness of student. In another study examining three flipped courses, Kim and colleagues (2014) found that students perceived classroom activities to be more student-oriented than teacher-oriented compared to activities in non-flipped classes and that peer interaction was helpful for their understanding of the class's core concepts.

It is important to identify through what mechanisms the flipped format has an effect on student outcomes. For example, Jensen and colleagues (2015) suggest that the benefits of the flipped format may be from students engaging in active learning; as such, they compared an active learning flipped class and an active learning non-flipped class. In the flipped class, content attainment (gaining a conceptual understanding of the material) took place before class and concept application (using the concepts in novel situations) took place during class, and in the non-flipped class, content attainment took place during class, and concept application took place after class. They found no statistically significant differences in achievement on unit exams, homework assignments, and final exam scores, suggesting that encouraging undergraduate instructors to use more active learning techniques might have the same benefits as flipping the classroom. A literature review from O'Flaherty and Phillips (2015) on 28 studies on the flipped classroom suggests that benefits found in the flipped are from indirect evidence (e.g., exam grades, perceptions, and staff satisfaction). Therefore, it is imperative for studies evaluating flipped instruction to include and test for additional measures that may explain why flipped instruction affect students' achievement.

The Flipped Classroom and Learning-Related Behaviors

There are a number of reasons why the flipped classroom can be beneficial for student learning. Theories that focus on the cognitive load students face during non-flipped courses could suggest that the flipped format allows students to ask questions more easily. Students do not have to keep track of points of confusion because class time is spent more actively rather than passively, allowing them to better learn the material. Therefore, students may adopt better study strategies such as keeping track of how they work and going back to check their answers. Students may also gain other skills from the flipped format. For example, Kong (2014) and colleagues looked at information literacy (gathering, synthesizing, interpreting, and evaluating information) and critical thinking skills (capability to think reflectively and judge skillfully) of students in a flipped Integrated Humanities class in a secondary school in Hong Kong. They found that their information literacy and critical thinking skills statistically significantly improved. Given the current evidence, research evaluating study skills gained from a flipped course may further the potential of the format helping improve STEM retention. Flipped classrooms might also benefit student learning through increasing student interest in the course such that students can see more relevance of the material during an interactive classroom session as compared with non-flipped instructor-centered approaches. In the current study, we explore differences in students' report of several learning-related outcomes between flipped and non-flipped instruction: study skills, attendance, and interest in the course. Although there is evidence of improved course grades and student satisfaction as a result of the flipped format, few studies are robust in looking at learning outcomes that follow students over time (see O'Flaherty & Phillips, 2015). Thus, there is a need for flipped classroom research to understand how students benefit from the format beyond the flipped course.

Study Designs for Evaluating Flipped Instruction

Differences in study designs in understanding the flipped classroom's effectiveness exist. For example, studies exploring student outcomes compare students in the flipped class to a previous course taught in a non-flipped format (Deslauriers et al., 2011; O'Flaherty & Phillips, 2015; Stelzer et al., 2010; Wilson, 2013). Stelzer and colleagues (2010) examined student grades from an introductory physics course from the Spring and Fall terms of 2008, and they compared them to the grades from the same course when it was taught from 1997 to 2002. However, this was a time difference of up to 10 years—student demographics, instructors, and exam content most likely have changed over time. Thus, each could have potentially played a factor in the differences found in outcomes between the flipped course and the non-flipped courses. Other studies on flipped instruction do not have another course as a comparison group (Roach, 2014; Wilson, 2014), thereby providing only limited evidence of validity of the study. In addition, another study design typical of evaluating the effectiveness of the flipped format is for instructors to implement the flipped format for a portion of the course during the term (Roach, 2014). Some even implement this for as little as three sessions of their entire course term (Moravec et al., 2010; Stelzer et al., 2010; Deslauriers et al., 2011). These three studies found that students in the flipped format did statistically significantly better on performance (measured by common exam score or course grade) compared to students who took the same course non-flipped. However, implementing the flipped format for only a portion of the course term makes it difficult to truly understand the associations between flipped instruction and student outcomes.

One way to further the research on the flipped format is to explore how students perform in a subsequent course. This is important to consider because student achievement in the current class could be confounded with course difficulty, teaching quality, and instructor grading leniency. Examining students' achievement in a subsequent course would allow researchers to better understand if the learning gains students experience in a flipped course transfer to the subsequent course. To our knowledge, there is only one study on the flipped format that looked at student performance in subsequent courses (Rias-Rohani & Walters, 2014). These students took a flipped engineering course and two subsequent non-flipped courses in a three-course series. Although students in the flipped course had statistically significantly higher grades than students in the non-flipped course, there were no statistically significant grade differences between these groups in both of the subsequent courses. Though this study was an important step in better understanding student performance in the subsequent course post-flipped, the design of the study compared students in the flipped condition to students taught by the same instructor in past non-flipped courses. As stated earlier, it is crucial for research on the flipped format to compare the treatment course to a concurrent control course where the control course is as comparable in difficulty, rigor, and teaching quality as possible. It is worth further exploring the skills students take away from the flipped class, especially as more schools are considering or are currently implementing the flipped format. If flipped classrooms do indeed help students learn content better, evaluating student achievement in a subsequent course will be an indicator of student learning of previous course material, especially in courses where material builds on itself.

In addition to understanding the benefits associated of the flipped classroom on student outcomes, it is also important to understand the context in which flipping the classroom can work. Exploring quality of instruction in large introductory STEM courses is worth considering for improving STEM enrollment and retention, yet few studies have explored the flip in large introductory STEM courses. One exception is Strayer (2012), who investigated student perceptions of flipped instruction in introductory statistics. Though the results showed that the students taking the course favored flipped instruction, Strayer (2012) recommended that perhaps the flipped format could be more suited for an upper-division course. Strayer noted that those in an upper-division flipped course may be more motivated than those in an introductory flipped course as they are taking a course specific to their major than a course that merely fulfills their degree requirements. This opens the question of whether the flipped format is a less suitable design for introductory courses.

Research investigating non-achievement outcomes such as student attendance and engagement have also been evaluated with similar study designs as those looking at student achievement outcomes. Specifically, some studies compared student perceptions in the flipped format to student perceptions in the same non-flipped course (Deslauriers et al., 2011; Mason et al., 2013), whereas other studies do not use a comparison group at all (Chen, Wang, Kinshuk, & Chen, 2014; Kim et al. 2014; Roach, 2014). For example, Deslauriers and colleagues (2011) compared student attendance of flipped and non-flipped courses conducted during the same term and found that student attendance increased significantly after a researcher came and taught the flipped version. It is difficult to know if student attendance increased as a result of the format or the new instructor. Likewise, studies using student surveys with low response rates (e.g., Kim et al., 2014), might limit the generalizability of the findings. In another study He and colleagues (2016) used data from students' self-reports of the amount of time they studied outside of class and found no differences between students in flipped and non-flipped courses. However, the authors noted that study time was measured with self-reports and were highly skewed. As such, it might not be the quantity, but the quality, of study time that may contribute to differences in achievement between students in flipped and non-flipped courses.

Table 1
Descriptive Statistics of Demographic Variables Across Sample 1 and 2

	Sample 1 (n=252)	Sample 2 (n=295)
Asian	45%	39%
Hispanic	29%	32%
White	9%	11%
Other ethnicity	17%	18%
Male	31%	33%
Low income	44%	49%
First generation student	65%	68%
STEM major	43%	44%
AP Chemistry exam	8%	9%

Note. Sample 1 is students who have valid survey data. Sample 2 is students who took the subsequent course in the sequence

The Present Study

The present study compared two sections of an undergraduate chemistry course—one non-flipped and the other flipped. Our study makes a unique contribution to the literature by investigating the associations between the course format and student learning experiences and outcomes, as well as comparing subsequent course performance of students who took a flipped course section and those who took the same course in a non-flipped section. We present the following research questions and hypotheses: (1) Do students in the flipped classroom report higher interest, use of study skills, and class attendance than students in the non-flipped classroom? Because the flipped format allows for active learning techniques in the classroom, we hypothesize that students enrolled in the flipped format will have higher interest, study skills, and attendance outcomes than students enrolled in the non-flipped format. (2) Do students enrolled in an undergraduate flipped format chemistry class earn higher grades in their subsequent chemistry course compared with students who enrolled in the non-flipped format? We hypothesize that students enrolled in the flipped format will earn higher grades in the subsequent course than students enrolled in the non-flipped format.

Method

Participants

The present study used data from a larger study investigating instructional practices in undergraduate STEM courses at a large public university in Southern California. The sample consisted of two sections of a large undergraduate introductory chemistry course taught in the Winter term of 2014 by two different instructors.

Six hundred and twelve students enrolled in Chemistry 1A (Chem1A) in the Winter term: 372 students enrolled in the flipped section, and 240 enrolled in the non-flipped section. Students in Chem1A had a mean SAT score of 592 (out of 800). Also, 44% were male, 38% were of Asian ethnicity, 32% Hispanic, 10% White, and 20% of other ethnicity (comprising of American Indian, Black, Pacific Islander, unknown, non-resident, or two or more ethnicities). Of these 612 students, 48% students subsequently enrolled in one of the two Chem1B sections (each taught by different instructors). Twenty percent of students who took the flipped Chem1A course were enrolled in the first listed Chem1B section (Tuesdays and Thursdays at 9:30 to 10:50am), whereas the remaining 80% were enrolled in the second listed Chem1B section (Mondays, Wednesdays, and Fridays at 12:00 to 12:50pm). Subsequently, 33% of students who took the non-flipped Chem1A course were enrolled in the first Chem1B section, while the remaining 67% were enrolled in the second section.

To answer our two research questions, we created two analysis samples: sample 1 consisted of students who participated in the surveys administered during their Chem1A course (n=252) and sample 2 consisted of students who continued on to the next course in the sequence (n=295). To be eligible in sample 1, students had to have valid responses on the pre and post surveys administered at the beginning and at the end of the Winter term. To be eligible for sample 2, students had to enroll and complete the next course in the sequence, Chem1B. Table 1 presents descriptive statistics for the overall sample and by course format for samples 1 and 2.

Regression analyses predicting completing the next course in the sequence (Chem1B) suggest that students in the flipped format course were not statistically significantly more likely to enroll in the subsequent course (p = .08) than students enrolled in the non-flipped format controlling for their Chem1A performance.

Context and Procedure

Instructors were recruited to participate in a larger study observing instructional practices in undergraduate STEM courses. As part of their participation in the study, instructors allowed researchers to administer two surveys—one at the beginning and one at the end of the term—to students for extra credit in the course. Surveys were administered via the university's online course management system, and students were able to access and complete the survey within one week after being made available. The response rate for the flipped section was 40% and for the non-flipped section 46%.

Both sections of Chem1A were part of that university's three-course introductory chemistry series—a mandatory prerequisite for Biology, Chemistry, Earth System Science, Public Health Science, Nursing Science and other related majors. Both sections had three one-hour sessions that took place on Mondays, Wednesdays, and Fridays for 10 weeks, one in the morning and the other in the early afternoon. Both sections had a mandatory formal class session component, which was led by the instructor of the section, and a discussion component, which was led by the teaching assistant of the section. Because this course is the first of a three-part introductory chemistry series, this class (Chem1A) is usually taken in the Fall term, the second introductory chemistry course (Chem1B) in the Winter term, and the third introductory course (Chem1C) in the Spring term. However, because the introductory chemistry course was offered in the Winter term, this class was likely to have a number of late-track students.

The flipped section had a class website for all announcements, video lecture materials, Powerpoint slides, homework links, office hours, and instructor and teaching assistant contact information. For each chapter, homework assignments counted as six percent of the student's total grade, while other homework assignments did not count towards the grade but were recommended to be completed. Videos were uploaded to YouTube with recommended assignments at the end of each video, and students were required to watch the videos before class (ranging from one to four videos per class). Videos were less than ten minutes in length and showed Powerpoint slides with audio spoken over them. In-class participation and quizzes (proctored at random on dates unknown to the students) contributed to three percent of the student's total grade. The instructor who taught the flipped section implemented the flipped method for the first time in the Fall 2014 but had been teaching the undergraduate introductory chemistry series at the university for over two years. A typical lesson in this flipped chemistry course varied in its degree of flipped implementation: there were some class periods where the instructor ran the class period flipped for the majority of class time and other class periods where the instructor lectured for the majority of class time. For example, in one one-hour class session toward the beginning of the term, the instructor asked her students to work on four multiple-part problems with each other from the beginning of the class time to nearly the end, and then the instructor instructed in the remaining time left. During these problem-solving sessions, students were able to have their questions answered by the teaching assistants and the instructor. In another class session toward the end of the term, the instructor used the majority of the class time to instruct, utilizing PowerPoint slides to present information to the class, and had students work on one problem collaboratively at the end of the class session time. Researchers were able to gather this information from observations, interviews with the instructor, and interviews with students taking the class.

The non-flipped section also had a course website consisting of announcements, instructor and teaching assistant contact information, office hours, lecture slides, discussion component worksheets and answer keys, weekly quiz answer keys, midterm answer keys, and final answer keys. The instructor of the course was a graduate student in his last year of his Chemistry Ph.D. Contrary to the flipped section, weekly quizzes were administered to students on Mondays the beginning of class, which counted for 20% of the grade. Instead of asking questions via email, students were encouraged to use a web platform called Piazza, where students ask questions for other students, teaching assistants, and the instructor to answer. The instructor used iClicker questions to gauge students' understanding and gave physical demonstrations of chemistry constructs during the lectures.

Measures

We collected data from two sources: Student surveys administered by the researchers twice during the term (Time 1 and Time 2) and student-level university data obtained from the Office of Institutional Research.

Interest. We measured interest for the course using three items from the Mathematics Value Inventory (Luttrell et al., 2010): "I find many topics in the course to be interesting", "Solving problems in this class is interesting for me", and "I find this class intellectually stimulating." Each item was on a scale of 1 to 5, 1 being "Never" and 5 being "Always". Students had to

Table 2

Descriptives of SAT Mathematics Scores, Learning Comparison of Interest and Study Skills, and Attendance (n = 252)

	Overall				Flipped (n=136)				Non-Flipped (n=116)			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Mathematics SAT Score	578.73	78.98	400	800	584.85	77.05	400	800	571.55	80.9	410	800
Time 1												
Interest	3.68	0.88	1	5	3.56*	0.83	1	5	3.80*	0.92	1	5
Study Skills	3.61	0.82	1	5	3.50*	0.77	1	5	3.74*	0.87	1	5
Time 2												
Interest	3.62	0.95	1.33	5	3.48**	0.91	1.33	5	3.80*	0.91	1.33	5
Study Skills	3.64	0.79	1.33	5	3.56	0.74	1.67	5	3.72	0.85	1.33	5
Attendance	4.67	0.62	2	5	4.75	0.55	2	5	4.57	0.68	2	5

Note. * $p < .05$ ** $p < .01$ *** $p < .001$. Asterisks indicate a significant difference between the means of the non-flipped and flipped groups using a two-sample T test.

have answered at least one of the items at the beginning and end of the term to be included in sample 1. Item reliability for the interest component was .83 at the pretest and .87 at the posttest. The mean interest at Time 1 was 3.68 with a standard deviation of 0.88 and ranged from 1 to 5 (see Table 2). The mean interest at Time 2 was 3.62 with a standard deviation of 0.95 and ranged from 1.33 to 5 (see Table 2).

Study skills. We used three items from the metacognitive strategies and practices from Wolters' adaptation of the Motivated Strategies for Learning Questionnaire (Wolters, 2004) to evaluate student learning behaviors. On a scale of 1 to 5, with 1 being "Not at all true" and 5 being "Very True," each item asked students to indicate how much they agreed with the following statements: "When I'm working, I stop once in awhile and go over what I have been doing", "Before starting an assignment, I try to figure out the best way to do it," and, "I keep track of how much I understand the work, not just if I am getting the right answers." We used the mean of the students' responses to the three items to measure their study skills at the beginning and the end of the term. Students had to have answered at least one of the items at the beginning and end of the term to be included in sample 1. Item reliability for student study skills was .75 at Time 1 and .75 at Time 2. The mean of study skills at Time 1 was 3.61 with a standard deviation of 0.82 and ranged from 1 to 5 (see Table 2). The mean of study skills at Time 2 was 3.64 with a standard deviation of 0.79 and a range of 1.33 to 5 (See Table 2).

Attendance. We used one item from the student survey administered at Time 2 to evaluate student self-reported attendance at the end of the term. We asked students to report on a scale of 1 to 5, from 1 as "Never" and 5 as "Always," to indicate how often students attended the class session. Mean attendance was 4.67 with a standard deviation of .0.62 and ranged from 2 to 5 (See Table 2).

Grade in subsequent course. Grades at the end of each course were collected from the university records. Grades were assessed on a 4-point scale, where 4.0 was an A, 3.7 was an A-, 3.3 was a B+, 3.0 was a B, 2.7 was a B-, 2.3 was a C+, 2.0 was a C, 1.7 was a C-, 1.3 was a D+, 1.0 was a D, and 0 was an F. This is a typical grading scale for large universities in the United States. For sample 2, the mean grade obtained in Chem 1B was 2.07 with a standard deviation of 1.21 and a range of 0 to 4 (See Table 3). Because we controlled for student grades in Chem1A, we also reported the mean (2.78), standard deviation (0.74), and range (1.7 to 4) on Table 3.

Covariates. University records provided to the research team included information on the students' ethnicity (Asian, Hispanic, White, and Other), SAT mathematics test score (continuous), whether the student met low-income level defined by Federal TRIO Program[1] (dichotomous), whether the student is a first generation college student (dichotomous), whether the student is a STEM major (dichotomous), and whether the student took the AP Chemistry exam in high school (dichotomous). In addition, we used the students' grade in Chem1A as a covariate for answering research question 2. The average grade in the flipped course was 2.82 (between a B- and a B) and the average grade for students in the non-flipped course was 2.72 (B-) on a scale of 1 to 4; differences between the average

[1] For more information on the Federal TRIO program see http://www2.ed.gov/about/offices/list/ope/trio/incomelevels.html

Table 3
Descriptive Grades and SAT Mathematics Scores for Sample 2

	Sample 2 (n=295)			
	Number	Chem1A Course Grade	Chem1B Course Grade	SAT Mathematics Score
		Overall		
Mean		2.78	2.07	579.05
SD		0.74	1.21	79.20
Min		1.7	0	400
Max		4	4	800
		Flipped		
Mean		2.82	2.45***	590.84**
SD		0.8	1.09	80.50
Min		1.7	0	400
Max		4	4	800
N	166			
		Traditional		
Mean		2.72	1.59***	563.88**
SD		0.65	1.2	75.11
Min		1.7	0	400
Max		4	4	800
N	129			

Note. ** $p < .01$ ***$p < .001$. Sample 1 is students who took the introductory chemistry course and received a grade. Sample 2 is students who took the subsequent course in the sequence. SD means standard deviation. Course grades are on a 4-point GPA score where a score of 4 is an A, 3.7 is an A-, 3.33 is a B+, 3.0 is a B, 2.7 is a B-, 2.3 is a C+, 2.0 is a C etc. Asterisks indicate a significant difference between the means of the traditional and flipped groups using a two-sample T test.

grades in both sections were not statistically significant (Table 3).

Analysis Plan

Research question 1. We asked: Do students in the flipped classroom report higher interest, use of study skills, and class attendance than students in the non-flipped classroom? All statistical analyses were performed with *Stata 13* (StataCorp, 2013). To predict whether the three outcomes were related to taking the course in a flipped format, we conducted three separate regression analyses on sample 1. For all models, except for student attendance, we controlled for students' previous (Time 1) reports of that construct. The analyses for this research question is based on the following Ordinary Least Squares (OLS) regression model:

$$YT2Survey_{iW} = \beta_0 + \beta_1 Flipped_{iW} + \beta_2 T1Survey_{iW} + \beta_3 Covariates_{iW} + e_i$$

where $YT2\ Survey_{iw}$ is the collection of interest, study skills, and attendance outcomes for student i, derived from their responses to our student survey at time 2; $Flipped_{iW}$ is a dichotomous indicator of the instructional format (flipped or non-flipped) of the student's chemistry course taken the Winter term and equals 1 if it was presented in a flipped format; $T1survey_{iw}$ is a continuous indicator of students responses at time 1 for the interest, study skills, and attendance outcomes, $Covariates_{iW}$ is the observed student characteristics described above for student i measured in the winter term; β_0 is a constant and e_i is an error term.

Research question 2. We also asked whether undergraduates who enrolled in a flipped format chemistry class earn higher grades in their subsequent chemistry course compared with those who enrolled in the non-flipped format. To predict the grade the student received in the subsequent course, we conducted an OLS regression analysis on sample 2. Our model is as follows:

(2) $Achievement_{is} = ßo + ß1Flipped + ß_2PriorAchievment_{iw} + ß3Covariates_{iw} + e_i$ where $Achievement_{iS}$ is a subsequent grade observed for student i in Chem 1B taken in the Spring term; $Flipped_{iW}$ is a dichotomous indicator of the instructional format (flipped or non-flipped) of the student's chemistry course taken the Winter term and equals 1 if it was presented in a flipped format; $PriorAchievment_{iw}$ is the grade observed for student i in Chem1A taken in the Winter term; $Covariates_{iW}$ are the observed student characteristics described above for student i measured in the Winter term; β_0 is a constant, and e_i is an error term. The subscript W refers to the Winter term, and the subscript S refers to the Spring term.

Results

Course Selection and Student SAT Score

Because the current study aimed to make comparisons between flipped versus non-flipped instruction, selection effects into the flipped classroom were of concern. To understand the extent to which this was a problem, we ran a series of logistic regressions in which class format was regressed on a series of demographic variables to understand pre-existing differences between the two groups. We tested whether gender, ethnicity, STEM major, first generation status, the taking of the AP exam in chemistry, low-income status, and initial SAT math were systematically related to students selecting the flipped format as opposed to the non-flipped format. We found evidence of selection effects for student scores on the SAT math exam where a one-unit increase in SAT math score is associated with an expected change in log odds for enrolling in the flipped section was .003 ($p = .002$). Because selection is always of concern in non-randomized trials, we use SAT math and other demographic characteristics as covariates.

Research Question 1

Table 4 presents the associations between the flipped format and interest, study skills, and attendance, while controlling for the covariates listed above. Model 1 presents the results predicting to their interest. We found that student interest in the flipped section did not significantly differ from interest in the non-flipped section at the end of the course controlling for initial interest ($\beta = -.16$, $p = .14$). Model 2 presents the results predicting to their study skills and we found no significant differences in their reported use of study skills by course format ($\beta = .00$, $p = .98$). Model 3 presents the results predicting to students' self-reported attendance. We found that students enrolled in the flipped section reported attending class more than students enrolled in the non-flipped format ($\beta = .32$, $p = .012$).

Research Question 2

Table 5 presents the results predicting students' grades earned in the next course in the sequence while controlling for our covariates. We found that on average, students enrolled in the flipped section of Chem1A obtained significantly higher grades in their Chem1B course than students who were enrolled in the non-flipped section ($\beta = .55$, $p < .001$) even after controlling for student-level characteristics. This can be interpreted as a .89 increase in student grade point average as measured on a 4-point scale—almost one full letter grade.

Validity Check

Because flipped instruction was confounded with instructor in our study, there were concerns about the validity of our findings. Specifically, it could be that students enrolled in the flipped section had higher achievement outcomes in the subsequent course because of an effective instructor and not due to the flipped format. To address this concern, we provide additional evidence about the instructors in the form of (1) comparing the syllabi of both professors and (2) using data available to us from the larger study so that we could see whether or not the flipped instructor had larger gains compared to other instructors more generally with other students during a different term. If the instructor is comparable to other instructors teaching Chem1A in the non-flipped format, it is more likely that the associations we found were due to the format and not just due to the instructor.

Comparing syllabi. We compared the syllabi of each course. According to each syllabus, the textbook for the course was the same for both courses (*Chemical Principles: The Quest for Insight* by Atkins, Jones, and Laverman, 6th edition). The flipped section facilitated in-class activities that were worth three percent of a student's total grade, whereas the non-flipped section facilitated in-class activities that were worth five percent of a student's total grade. Both courses worked through example problems in lecture. Both sections assigned graded homework assignments on Sapling Learning (Sapling Learning, 2011), an online software that provides interactive learning experiences. Assigned homework in the flipped section was worth six percent of the grade, and in the non-flipped it was worth 10% percent. Both classes also had non-graded homework problems from the same textbook, and both syllabi recommended these problems as good practice. Though the flipped section's midterms were each worth 25% of the total grade and the final was worth 38% while the non-flipped section's midterms were worth 20% and the final exam was worth 25%, both courses

Table 4
Associations between Flipped Format and Interest, Study Skills, and Attendance (n=252)

	Interest	Study Skills	Attendance
Flipped	-0.16	0.00	0.32*
	(0.11)	(0.11)	(0.13)
SAT mathematics	-0.10	-0.04	-0.03
	(0.06)	(0.06)	(0.07)
AP Chemistry	0.08	0.02	0.02
	(0.20)	(0.20)	(0.23)
Male	-0.11	-0.02	-0.14
	(0.12)	(0.12)	(0.14)
Hispanic	0.07	0.21	0.06
	(0.14)	(0.14)	(0.17)
White	0.02	0.26	0.34
	(0.20)	(0.20)	(0.23)
Other ethnicity	0.16	0.20	-0.02
	(0.15)	(0.15)	(0.18)
Low income	-0.02	-0.01	0.20
	(0.12)	(0.12)	(0.14)
First generation	0.04	0.06	0.16
	(0.14)	(0.13)	(0.16)
STEM major	-0.02	-0.04	-0.15
	(0.11)	(0.11)	(0.13)
T1 survey	0.51***	0.54***	
	(0.06)	(0.05)	
Constant	0.06	-0.13	-0.31
	(0.15)	(0.15)	(0.17)
R-sq	0.314	0.313	0.072

Note. * $p < .05$ ** $p < .01$ *** $p < .001$. Standard errors in parentheses. All continuous variables are standardized. The reference group is Non-Flipped, Female, Asian, no AP Chemistry exam, not low income, not first generation, not STEM major. T1 survey refers to students' report of their interest and study skills at the beginning of the course. Information on students' attendance at the beginning of the term was not available.

administered the same number of exams: two midterms and a final.

Comparing the flipped instructor to other chemistry instructors. To alleviate some concern over the issue of teaching quality confounding the results of our study, we were able to utilize data collected as part of the larger study from instructors teaching the same course (Chem1A) to on-track students in the Fall term. The instructor of the flipped section in the Winter taught a flipped version of the course in the previous Fall. To understand whether or not the instructor who taught the flipped section was higher in teacher quality overall, as measured by gains in student achievement in the subsequent course, we conducted OLS regression to compare gains in achievement elicited by the flipped instructor to gains in achievement elicited by the three other instructors. Using the same covariates to answer research question two, we found that students who took the non-flipped on-track Chem1A course in the Fall by the flipped instructor did not perform statistically significantly differently from students in the Chem1B course taught by other instructors during the same term ($\beta = .05$, *ns*; see Table A in the appendix). This provides some evidence to suggest that the teaching quality of the flipped instructor was not statistically significantly higher than other professors at that university.[2]

[2] We were unable to follow this procedure to understand if the instructor of the non-flipped course was lower in overall teaching quality because the research team did not collect additional data on other courses taught that academic year by the non-flipped instructor.

Table 5
Association Between Flipped Format and Subsequent Course Grade (n=295)

	Subsequent Course Grade
Flipped	0.55***
	(0.08)
Chem1A grade	0.55***
	(0.04)
SAT mathematics	0.24***
	(0.05)
AP Chemistry	0.25
	(0.14)
Male	0.12
	(0.09)
Hispanic	0.17
	(0.10)
White	0.38**
	(0.13)
Other ethnicity	0.16
	(0.11)
Low income	-0.01
	(0.08)
First generation	-0.01
	(0.10)
STEM major	-0.08
	(0.08)
Constant	-0.45***
	(0.11)
R-sq	0.572

Note. * $p < .05$ ** $p < .01$ *** $p < .001$. Standard errors in parentheses. All continuous variables are standardized. The reference group is Non-Flipped, Asian, Female, no AP Chemistry exam, not low income, not first generation, not STEM major.

Discussion

The present study investigated whether students in the flipped format reported higher interest, use of study skills, and class attendance compared to students in the non-flipped format. The present study also examined whether students in a flipped introductory chemistry course earned higher grades in the subsequent chemistry course than students in a non-flipped chemistry course. For study skills and interest, no statistically significant differences were found between students in both courses; however, we did find that students in the flipped section reported higher class attendance, which is consistent with previous findings (see O'Flaherty & Phillips, 2015). We also found that students who took the prior course in the flipped format did statistically significantly better in the subsequent course than students who took the prior course in the non-flipped format suggesting that the effectiveness of flipped instruction may extend beyond that current course.

The present study builds upon and extends the literature in significant ways. Because flipped instruction introduces more active learning elements in the classroom, we posited that students would report on more adaptive study skills, higher interest for the course, and higher class attendance than students in the non-flipped class. We only found evidence suggesting that students reported attending class more in the flipped section than in the non-flipped class. Though the format of the flipped classroom would suggest that students may be better able to capitalize on more effective study skills such as keeping track of what a student is learning, we do not find evidence supporting this claim. Likewise, because flipped classrooms use more active learning approaches, we hypothesized that students would report higher interest in the course in the flipped format compared to the non-flipped format. However, because the non-flipped format also made use of active learning strategies such as the use of iClickers, it may be that active learning strategies in general pique the interest of students and encourage them to use more adaptive strategies and not the flipped format. Since we did find statistically significant differences in students' grades in the subsequent course in the sequence, we suggest that other mechanisms such as quality of information retained or continual use of adaptive study strategies may explain our results.

Our study builds on the methodological limitations of previous work that have used comparison groups from different terms. Though there are studies on the flipped format that have compared performance to that of a concurrent control course (Deslauriers & Wieman, 2011; Mason et al., 2013; Strayer, 2012) and a study that looks at performance of students who took a flipped course in the subsequent courses (Rais-Rohani & Walters, 2014), our study is the first to do both. This adds to the suggestion that the students who have taken the prior course flipped can earn higher grades in the subsequent course compared to students who, at the same time, took the prior course non-flipped.

Additionally, our study further suggests that the flipped format may work well in a large introductory course (Deslauriers et al., 2011; Moravec et al., 2010). Whereas Strayer (2012) suggested that the flipped format might be better suited for upper-division courses, our study highlights the potential for the flipped format to work in introductory courses. Similar to Moravec and colleagues (2010), who found significantly higher student performance in a large flipped introductory biology course compared to performance in large, non-flipped biology courses, the findings of the present study highlight the potential of the flipped format in large introductory courses. Further studies should be conducted to directly compare student outcomes as a result of the flipped format in upper-division versus lower-division university courses.

Another way this study builds upon the current literature is that it is one of the few studies to investigate the flipped format on students in different academic tracks. One exception is Morin and colleagues (2013) who found no significant differences in performance of students taking an honors flipped course compared with students taking a non-flipped version of the same course. However, students in the flipped course were enrolled during the first term, whereas the comparison group (non-flipped) enrolled during a different term than when the honors course was normally taught. As Morin and colleagues (2013) noted, it could be that students who took the non-flipped course could have been dismissed from the honors program or could have been taking the course late track, suggesting that the groups were not comparable. The results of the present study provide the potential for the flipped format to be effective in late-track classes; however, it is still unclear whether the flipped format works for all late-track students or if there are differential effects of the flipped format on students of different ability levels.

Limitations and Future Directions

We note several limitations of this work. Due to the number of courses taught in the flipped format, our sample only consisted of students from two courses. These results from the study should be interpreted cautiously because instructors were not randomly assigned to teach in the flipped format, and students were not randomly assigned to course format.

Instructor-level characteristics that might be related to whether they decide to teach in a particular format could not be investigated in our study. However, in practice, this may not be a grave limitation. Though it is important to use random assignment to further understand the effects of the flipped format, because instructors usually decide whether or not to flip their course, we believe it is essential to first study instructors who choose to use the flipped pedagogy. In the present study, we used data from two different instructors teaching the same chemistry course. Though differences in students' outcomes could be attributed to the quality of the instructor teaching the flipped course (recall, the instructor of the non-flipped course was a graduate student), we were able to capitalize on data from the larger study and compare student outcomes from the flipped instructor teaching a different course in a different term, with other instructors teaching that same course. We found no statistically significant difference in students' outcomes when comparing the flipped instructor with other instructors, suggesting that the flipped instructor was not just generally a better instructor. Though this validity check gives us some confidence in our findings that it was indeed the flipped format that lead to our results, we were limited in that we were unable to conduct the same such validity check with the non-flipped instructor.

In addition to the lack of random assignment of instructors, there was no random assignment of students. As mentioned in our results section, we found a small but statistically significant difference in students' SAT math scores between the two Chem1A sections. We found further evidence of this in our informal interviews we conducted with students in the flipped section; some reported that they enrolled in the course because they knew it was flipped. However, other students we interviewed were not aware of this, and to our knowledge the course was not advertised as a flipped course. Most students from both groups who completed Chem1B enrolled in the second section, which further supports the lack of random assignment limitation. This opens up the question of what are the specific characteristics of students who enroll in the flipped course? Do they tend to be high-achieving, more motivated, and/or more conscientious? Future research could explore the characteristics of students who enroll in the flipped format and whether the format is particularly effective for students who have certain characteristics.

It is possible that our measures of interest and study skills were not reliably or validly measured. Though the reliabilities for the measures were within the range of what is commonly accepted (alpha coefficients ranging from .75 to .87), using too few items to measure a construct could influence the validity of the findings such that we may not have measured the breadth of the constructs. Due to time constraints on the student survey, the research team could only include a small number of items for each construct that was measured. As such, we urge future research to replicate these findings using more complete survey measures. It may also be that students study skills do not change that quickly, and instead, an effect on students' study skills in the subsequent quarter should be investigated. Unfortunately, due to the timing of the larger study, we were unable to do so.

We note the limitation of ceiling effects in our measure of self-reported attendance. While we found significant differences between students enrolled in the flipped format and students enrolled in the non-flipped format, the mean of self-reported attendance was 4.67 on a 5-point scale. Because students are likely to report that they attended classes more than they actually did, future studies could explore more objective ways of measuring attendance such as through the use of observations that do not rely on self-report. Likewise, our study also had a low response rate to the survey even though students were incentivized to take the surveys for extra credit in the course. We wondered whether the students who took the survey were highly motivated/conscientious or if they were the ones most in need of extra credit. Though we do not have survey data on motivation outcomes, to explore this with the data given, we ran a correlation to see whether taking the survey was related to students' previous achievement as measured by their mathematics SAT score. We found that taking the survey was negatively correlated with mathematics SAT score ($r = .08$, $p < .001$), and if we consider their previous achievement as a proxy for motivation, perhaps students less motivated/conscientious took the survey because they were in need of extra credit. Future research can explore whether students who take the survey are more or less motivated/conscientious.

Though self-reported data is a limitation, we believe our self-reported attendance data serves as a starting point for future research. As participation in in-class activities were weighted similarly across both course sections (3% for the flipped course and 5% for the non-flipped course), this suggests that differences in the grading of class participation may not be a factor in whether a student decides to attend class. In-class quizzes in the flipped course were worth three percent of the student's total grade but were pop-quizzes (students did not know ahead of time when the quizzes would be administered), and in-class quizzes in the non-flipped were worth 20% of the student's total grade but were not pop-quizzes (students knew ahead of time when the quizzes would be administered). It is difficult to tell with the available information whether the pop quizzes in the flipped course were the drivers of student attendance, but since quizzes in the flipped course were only worth three percent of a student's total grade (as opposed to 20%), it may be likely that other factors contributed to students reporting higher attendance in the flipped course versus the non-flipped course. In regards to the issue of self-reported attendance as being a proxy for student engagement, we do not contest this assertion. It could very well be that students who attended class more were the ones who were more likely to engage in class; however, without additional data such as interviews from students, we are unable to know. As such, self-reported attendance could be an indicator of student engagement. Future research could explore whether student engagement mediates attendance rates in the flipped format. The flipped classroom was still relatively new to the university and was the instructor's first time teaching the course flipped. Therefore, the implementation of the flipped design may not have been consistent throughout the term. From informal interviews and observations, the entire class period was sometimes not dedicated to working through problems and collaborative learning. Therefore, further research could investigate whether or not different degrees of flipped instruction are most adaptive for student outcomes, perhaps through developing some measure that evaluates the extent of which the course is flipped.

Conclusion

As active learning designs have been proposed as a strategy that could potentially increase the retention rate of STEM majors, our study provides a closer look into this specific approach of active learning. By exploring student achievement and learning-related behaviors in the flipped classroom, we hope our work encourages efforts to increase STEM retention rates across universities such that students will graduate with their degrees and be prepared for technology- and information-based careers.

References

Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers & Education, 78,* 227-236. doi: 10.1016/j.compedu.2014.06.006http://dx.doi.org/10.1016/j.compedu.2014.06.006

Baillie, C., & Fitzgerald, G. (2010). Motivation and attrition in engineering students. *European Journal of Engineering Education*, *25*(2), 145-155. doi: 10.1080/0304379003088544

Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D. L., Allain, R. J., Bonham, S.W., Dancy, M. H., & Risley, J. D. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-based Reform of University Physics*, *1*(1), 2-39.

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day.* Alexandria, VA: International Society for Technology in Education.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *In ASEE National Conference Proceedings*, Atlanta, GA.

Chen, Y., Wang, Y., Kinshuk, & Chen, N. (2014). Is FLIP enough? Or should we use the FLIPPED model instead? *Computers & Education*, *79*, 16-27.

Clark, K. R. (2015). The effects of the flipped model of instruction on student engagement and performance in the secondary mathematics classroom. *Journal of Educators Online*, *12*(1), 91-115.

Consortium for Student Retention Data Exchange. (2013). *STEM retention report*. University of Oklahoma Outreach. Retrieved from http://csrde.ou.edu/web/Samples_Web/Sample-STEMReport.pdf

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.

Fairweather, J. (2008). Linking evidence and promising practice in science, technology, engineering, and mathematics (STEM) undergraduate education. *Board of Science Education, National Research Council, The National Academies, Washington, DC.* Retrieved from http://otl.wayne.edu/wider/linking_evidence--fairweather.pdf

Freeman, S., Eddy, S., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wendroth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, USA*, *111*, 8410-8415.

He, W., Holton, A., Farkas, G., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction*, *45*, 61-71.

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

Khousmi, A., & Hadjou, B. (2005). Learning probabilities in computer engineering using a competency – and problem-based approach. *Journal of STEM Education, 6,* 5-13.

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (July 2014). The experience of three flipped classrooms in an urban university: an exploration of design principles. *The Internet and Higher Education, 22,* 37-50.

Kong, S. C. (2014). Developing information literacy and critical thinking skills through domain knowledge learning in digital classrooms: An experience of practicing flipped classroom strategy. *Computers & Education*, *78*, 160-173.

Kyle, W. C. (1997). Editorial: The imperative to improve undergraduate education in science, mathematics, engineering, and technology. *Journal of Research in Science Teaching*, *34*(6), 547-549.

Luttrell, V. R., Callen, B.W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. (2010). The mathematics value inventory for general education students: Development and initial validation. *Educational and Psychological Measurement*. *70*(1), 142-160. doi:10.1177/0013164409344526

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 Education, 56*(4), 430-435.

McCallum, S., Schultz, J., Selke, K., & Spartz, J. (2015). An examination of the flipped classroom approach on college student academic involvement. *International Journal of teaching and Learning in Higher Education*, *27*(1), 42-55.

McGinn, M. K., & Roth, W. M. (1999). Preparing students for competent scientific practice: Implications of recent research in science and technology studies. *Educational Researcher*, *28*(3), 14-24.

Mervis, J. (2010). A way to heal science education, but is there the political will? *Science*, *329*(5999), 1582-1583.

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.

Morin, B., Kecskemety, K. M., Harper, K. A., & Clingan, P. A. (2013). The inverted classroom in a first-year engineering course. *In 120th ASEE Annual Conference & Exposition: Frankly We Do Give a D*mn.* Atlanta, GA.

Muzyk, A., Fuller, S., Jiroutek, M. R., Grochowski, C. O. C., Butler, A. C., & May, D. B. (2015). Implementation of a flipped classroom model to teach psychopharmacotherapy to third-year Doctor of Pharmacy (PharmD) students. *Pharmacy Education*, *15*(1), 44-53.

National Academy of Engineering. (2005). *NAE Annual Report*. Washington, DC: Author. Retrieved from https://www.nae.edu/File.aspx?id=43367

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

Rais-Rohani, M., & Walters, A. (2014). Preliminary assessment of the emporium model in a redesigned engineering mechanics course. *Advances in Engineering Education, 4*(1) 1-19.

Roach, T. (2014). Student perceptions toward flipped learning: New methods to increase interaction and active learning in economics. *International Review of Economics Education*, *17*, 74-84.

Sapling Learning. (2011). *Sapling Learning*. Austin, TX: Macmillian Learning.

StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

Stelzer, T., Brookes, D. T., Gladding, G., & Mestre, J. P. (2010). Impact of multimedia learning modules on an introductory course on electricity and magnetism. *American Journal of Physics, 78*(7), 755-759.

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research, 15*(2), 171-193.

Wolters, C. A. (2004). Advancing achievement goal theory: using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology*, *96*(2), 236-250.

Wilson, S. (2013). The flipped class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology*, *40*(3). 193-199.

TIMOTHY CASASOLA is an Organizational Design Consultant at The Ready, a New York City-based global consulting firm. Prior, he was an honors researcher and undergraduate at the University of California, Irvine, where his research focused on industrial-organizational psychology, facial expressions and student evaluations, and the flipped classroom.

KATERINA SCHENKE is a postdoctoral scholar at the Graduate School of Education and Information Studies at the University of California, Los Angeles. She studies teacher effectiveness, student motivation, and large-scale educational assessment. Her work has been published in journals such as *Computers and Education*, *Journal of Educational Psychology*, and *Contemporary Educational Psychology*.

TUTRANG NGUYEN is a doctoral student at the University of California, Irvine. Her research focuses on understanding individual and environmental factors that lead to healthy child development. Her work has been published in Early Childhood Research Quarterly and Developmental Psychology.

MARK WARSCHAUER is Professor of Education at the University of California, Irvine, where he directs both the Digital Learning Lab and the Teaching and Learning Research Center. Warschauer is a fellow of the American Educational Research Association and editor-in-chief of *AERA Open.* His books include *Learning in the Cloud: How (and Why) to Transform Schools with Digital Media* (Teachers College Press), *Laptops and Literacy: Learning in the Wireless Classroom* (Teachers College Press), and *Technology and Social Inclusion: Rethinking the Digital Divide* (MIT Press).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants 1256500 and 1535300, as well as the Undergraduate Research Opportunities Program and the Summer Undergraduate Research Program at the University of California Irvine. We want to thank Arena C. Lam, Kenneth Lee, Anna-Lena Dicke, Amanda Acevedo, Tyler Watts, and Cathy Tran for providing feedback and comments on previous drafts of this paper. We thank Dr. Anne McDaniel and Dr. Thurston Domina for being principal investigators of the larger study and Lynn C. Reimer and Maureen Spanier for participating in the data collection of the larger project. Lastly, we thank the instructors and students for participating in this study.

Appendix

Table A

Validity Check: Association between Fall Instructor and Subsequent Course Grade (n=1072)

	Subsequent Course Grade (Winter 2014)
Flipped Instructor	0.05
	(0.05)
Fall Chem1A grade	0.59***
	(0.02)
SAT mathematics	0.19***
	(0.03)
AP Chemistry	0.13*
	(0.05)
Gender	
Male	-0.03
	(0.05)
Ethnicity	
Hispanic	0.00
	(0.06)
White	-0.01
	(0.06)
Other	0.16*
	(0.07)
Low income	-0.00
	(0.05)
First generation	-0.13**
	(0.05)
STEM major	-0.04
	(0.05)
Constant	0.06
	(0.06)

Note. * $p < 0.05$ ** $p < .01$ *** $p < .001$. Standard errors in parentheses. All continuous variables are standardized. The reference group is Non-Flipped Instructor, Asian, Female, no AP Chemistry exam, not low income, not first generation, not STEM major.