JOURNAL OF CHEMICAL EDUCATION

Article

Cite This: J. Chem. Educ. XXXX, XXX, XXX−XXX

pubs.acs.org/jchemeduc

Dissecting the Flipped Classroom: Using a Randomized Controlled Trial Experiment to Determine When Student Learning Occurs

Matthew D. Casselman,*[†,§] Kinnari Atit,[‡,§] Grace Henbest,[†] Cybill Guregyan,[†] Kiana Mortezaei,[†] and Jack F. Eichler*[†,§]

[†]Department of Chemistry, University of California, Riverside, 501 Big Springs Road, Riverside, California 92521, United States

[‡]Graduate School of Education, University of California, Riverside, 1207 Sproul Hall, Riverside, California 92521, United States

Supporting Information

ABSTRACT: The use of the flipped classroom approach in higher education STEM courses has rapidly increased over the past decade, and it appears this type of learning environment will play an important role in improving student success and retention in undergraduate chemistry "gatekeeper" courses. Many adopters of the flipped classroom structure see the greatest benefit originating from the additional time this format provides for the implementation of student-centered learning activities during the classroom period. However, results from recent quasi-experiments suggest that improved course performance for students in flipped classroom environments has a significant contribution from the online preclass activities. In order to compare the impact of the preclass online learning environment to the in-class collaborative activities typically done in a flipped classroom, a randomized controlled trial (RCT) was conducted with student volunteers. A two-day organic chemistry stereochemistry unit was delivered to students who were randomly assigned to "flipped classroom" and "traditional lecture" treatment groups. Performance gains were measured after each phase of the instructional intervention for both treatment groups, and these gains were compared to students from a randomly assigned negative control group. A mixed-methods ANOVA indicates that under these experimental conditions the online learning component appears to account for most of the improvement in posttest scores observed in the flipped classroom treatment. These results suggest optimizing the design of the asynchronous online learning environment will positively impact student performance outcomes. Therefore, this component of the flipped classroom deserves more attention from instructional designers and classroom practitioners.

KEYWORDS: *Chemical Education Research, First-Year Undergraduate/General, Multimedia-Based Learning, Collaborative/Cooperative Learning, Stereochemistry*

FEATURE: Chemical Education Research

■ INTRODUCTION

Flipped classrooms and blended learning environments have been rapidly growing in popularity in the past 10 years, particularly in higher education STEM courses. This is reflected in the educational research literature, where the number of publications dedicated to flipped classrooms has grown significantly as online learning has become increasingly normalized in both secondary schools and in higher education (see Figure 1). Though this approach to teaching has been gaining momentum, there are still very few research publications dedicated to determining which aspect of the flipped classroom has the most significant impact on student classroom performance. In particular, it would be of interest to determine how much improvement on in-class assessments can be attributed to the online learning environment versus the in-class activities, and not just the overall flipped classroom treatment.

The exact definition of the flipped classroom can vary depending on the audience or discipline in which it is being discussed. Therefore, for the purposes of this study, flipped classrooms will be defined as a blended learning scaffold in which some portion of the traditional in-class lecture is completed by students as preclass work in an online, asynchronous format, and some portion of the traditional individual homework is done via collaborative group work that is completed in the live classroom. A more detailed discussion of the typical definition of the flipped classroom and what elements comprise this type of classroom structure has been previously reported by Seery.[1] We note here that a "flipped class" specifically refers to inverting the location and support structure of the lecture and homework components, with the lecture component taking place online and being completed individually, and the homework being done collaboratively in groups during the in-person classroom meeting. Some of the seminal work on flipped classrooms in the physics community associated the primary benefit of the intervention with the increased use of active learning techniques during class,[2,3] and a more recent review of flipped classroom implementations in undergraduate chemistry also finds that many educators view this classroom structure as a means to develop a more active learning environment and increase student engagement.[1] Recognition of this benefit is not surprising, given the fact the education research literature clearly indicates active learning pedagogies lead to greater learning gains than

Received: August 17, 2019
Revised: October 21, 2019

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

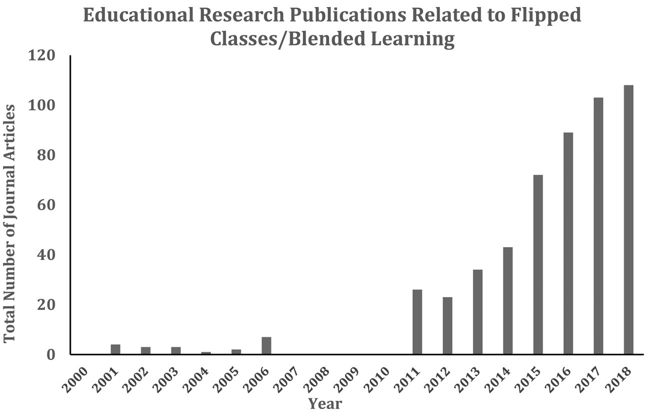

Figure 1. Total number of citations for educational research articles related to flipped classrooms plotted versus year (2000−2018). The total number of citations/year includes only research articles identified by the Web of Science database associated with the search terms "flipped classroom" and/or "hybrid learning".

traditional passive lectures.[4] However, a closer examination of the learning theory suggests the online asynchronous activities should also have a significant impact on students' learning outcomes.

Though an overview describing how cognitive load theory can act as the theoretical framework for the flipped classroom has been previously laid out by Seery,[1] additional context about the theory underpinning this type of learning environment is provided here. To explain the potential benefits and practicalities of implementing a flipped classroom pedagogical approach, we feel it is prudent to consult John Sweller's cognitive load theory[5] and the theory of knowledge coconstruction presented by Hogan.[6] Cognitive load theory asserts that because one's working memory capacity and short-term memory are limited, learning experiences ought to be designed in a way that allows students to more effectively learn new information and integrate it into their existing knowledge state. The flipped classroom approach recognizes this by replacing the rigidly structured traditional classroom environment with flexible preclass work that provides novice learners the necessary time to process and understand the information at their own pace. Indeed, on the basis of Seery's survey of flipped classrooms in higher education chemistry courses, a general theme emerged in which students reported liking this classroom structure because they can access and engage with the material on their own time.[1] Hogan would posit that students are then able to engage actively in collaborative work during class instructional time with problems that would otherwise be done as individual out-of-class homework. By moving the traditional individual homework to this collaborative setting, students acquire intellectual skills such as coconstructing and enriching knowledge, coupled with the ability to reason critically, that are fundamental in science learning and education. It seems evident that by adopting a flipped classroom model, STEM educators can use individually navigated online learning experiences to overcome the working memory and cognitive load limitations inherent in traditional lectures, and use the in-class synchronous learning to create a collaborative learning environment that will more effectively integrate this new knowledge into the students' long-term memory and existing mental models.

The concepts highlighted here regarding Sweller's cognitive load theory also constitute the foundational underpinnings of Richard Mayer's cognitive theory of multimedia learning.[7] The cognitive theory of multimedia learning purports that gaining new knowledge through online instruction can effectively take place if one receives information from two sensory platforms: auditory and visual. This information is then organized and works in conjunction with prior knowledge that is retrieved from one's long-term memory to foster meaningful learning as opposed to rote learning and helps overcome the limitations of one's working memory capacity.[8] The online preclass assignments used in this study were informed by the cognitive theory of multimedia learning, as they were structured in a way that integrated the auditory and visual components necessary for learning novel skills and new concepts.

The number of studies aiming to measure the efficacy of the flipped classroom in undergraduate education courses has increased rapidly over recent years. Seery's review of the flipped classroom summarizes the general trends about how instructors have designed flipped classrooms in chemistry and the efficacy of this classroom structure on student learning outcomes.[1] This analysis reveals that flipped classrooms generally lead to improved student outcomes, either in regard to improved exam scores, improved course grades, and/or reduced failure rates. We also highlight here two more recent quasi-experimental studies in which flipped classroom interventions were compared to "teaching as usual" controls and analyses probing the impact on less-prepared students were conducted. Ryan and Reid ran a parallel controlled study in general chemistry and found exam performance improved in the flipped treatment condition relative to a traditional lecture condition, but only for the bottom one-third of students based on academic preparation. The flipped treatment also lowered the DFW (D's, F's, and withdrawals) rate by 56%.[9] Crimmins and Midkiff compared student learning in a flipped classroom in organic chemistry to historical data from traditional lectures and found that all students in the flipped classroom scored higher on the exams and had higher probabilities of increased course grades relative to students from previous traditional lecture courses.[10] Notably, it was also found that students from the lowest academic achievement levels made the most gains. These findings not only corroborate the previous consensus that flipped classroom structures generally improve student performance in both general and organic chemistry, but they have also begun to provide more insight about which types of students might benefit most from this learning environment.

As mentioned above, many practitioners and researchers regard the in-class active learning as the most important aspect of the flipped classroom when it comes to improving student performance outcomes. However, recent quasi-experimental studies suggest the online learning component of the flipped classroom model may indeed account for a significant portion of the improved scores on in-class assessments and/or course grades. A previous report from our group revealed that student grades in a flipped classroom were significantly higher compared to a control section in which a more traditional lecture was administered.[11] Our initial motivation for investigating the efficacy of the flipped classroom also originated from the perceived benefits of incorporating more active learning in class. Further analysis of student performance on exams and in-class clicker questions indicated that, even though the flipped classroom and traditional lecture cohorts had equivalent exam scores, the students in the flipped classroom treatment scored significantly higher on the in-class clicker questions. These results suggested students in the

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 1. Research Study Design with Treatment Group and Control Group Activities over the Seven Day Study

Day	Full Flip Treatment (n = 15)	Traditional Lecture Treatment (n = 18)	Negative Control (n = 16)
Day 0	Pretest; spatial reasoning test	Pretest; spatial reasoning test	Pretest; spatial reasoning test
Day 4	Video lecture with Playposit[a] questions and online quiz (1.5 h)	In-person traditional lecture (1.5 h)	No lesson
Day 5	Posttest 1; Collaborative group learning/practice problems (1.5 h)	Posttest 1; Individual homework problems (1.5 h)	No lesson
Day 7	Posttest 2	Posttest 2	Posttest 2
Two weeks subsequent to study	Follow-up interviews with select participants	Follow-up interviews with select participants	n/a

[a]Playposit questions are questions embedded within the video lecture; see ref 17.

flipped classroom treatment initially performed better in class, presumably due to the preclass online learning activities, yet the in-class activities and homework completed in the traditional lecture control allowed these students to make equivalent learning gains on the summative assessments. A more recent quasi-experimental study compared a course taught using a flipped format to both a traditional lecture and a course that integrated collaborative active learning without any preclass learning.[12] This study found the flipped classroom that included both online learning and in-person collaborative active learning led to higher exam scores relative to the traditional lecture course and the course that included only a collaborative active learning intervention. These results further corroborate that online learning may be a more critical component of the flipped classroom than generally recognized.

Due to the restraints inherent in carrying out quasi-experimental studies within the confines of live classrooms, a randomized controlled trial (RCT) was employed outside of an actual classroom setting in the present study. In particular, this provided an opportunity to measure the impact of the different components of the flipped classroom on student performance without raising ethical questions regarding delivering equally effective instruction to all study participants and helped eliminate potential compensatory rivalry validity threats (i.e., when a study group not receiving a treatment feels disadvantaged). A mixed-effects research study design was used in which two alternative treatments, a flipped classroom and traditional lecture, were monitored for performance gains over time and compared to a negative control group in which no treatment was delivered. The RCT experiment provided an opportunity to test the specific research hypothesis stating, "The pre-class online learning component of a flipped classroom contributes as much, or more than the in-person collaborative group learning to the overall performance gains in a model flipped classroom intervention." The quantitative study was supplemented with follow-up interviews conducted with a subpopulation of participants from each treatment group, which provided additional insight about which aspects of the flipped classroom were most positively viewed by students and why they were motivated to participate. Hence, the results of this novel RCT experiment and qualitative analysis of the student experience in a flipped classroom environment will be described herein.

■ RESEARCH QUESTIONS AND METHODS

Research Study Design

This study was intended to test the following research questions: (1) How much of the student performance gains in a model flipped classroom intervention can be attributed to the preclass online learning? (2) How would the sequence of preclass online learning and in-person collaborative group problem solving impact student performance relative to the sequence of in-person passive lecture and individual homework problem solving? It is noted the primary motivation for including a traditional lecture treatment group was not to compare the pre/posttest performance gains between the flipped and traditional lecture group, but rather to provide a point of comparison regarding when student scores on the posttest assessments improved over the course of the study.

An RCT was conducted with student volunteers (n = 49) recruited from a first-quarter organic chemistry course taught in the fall of 2018. This course routinely incorporates collaborative group learning in class; therefore, all study participants had engaged in such activities prior to the RCT experiment. The sample size required for an analysis of variance (ANOVA) to achieve a power of 0.80 with an effect size of 0.50 was estimated as described by Cohen (three study groups with an α = 0.05; minimum number of participants per group = 14).[13] Student volunteers were recruited by the instructor via email, under the University of California-Riverside Institutional Research Board (IRB) human subjects protocol HS-16-212, and the recruitment clearly stated the study was not part of the official course. Students were asked to provide their availability in the recruitment email, and the final list of study participants was chosen on the basis of who was able to attend the scheduled study sessions.

The available participants were then randomly assigned to one of the three study groups prior to commencing the study (see Table 1), and were subsequently informed that their level of participation would have no bearing on their class grade and that they would be receiving a $25 gift card at the conclusion of the study. Students randomly assigned to the negative control group only completed the day 0 and day 7 pre- and posttests and did not receive any learning interventions. These students provided a baseline comparison for the performance gains measured in the two treatment groups. The flipped classroom and traditional lecture treatment groups attended two learning intervention sessions during the seven-day study period during the first and second weeks of the term where they learned topics relating to stereochemistry, and the sessions were held in the evenings outside of the time frame in which most classes on campus are scheduled. The study was carefully timed within the academic term such that students would have had an introduction to organic chemistry, specifically line-bond notation, but had not yet been introduced to stereochemistry as a concept. In the first session (day 0), all participants completed the pretest assessment and a measure of spatial reasoning (the Mental Rotations Test; MRT).[14] As stereochemistry is a spatially demanding topic in chemistry,[15] we measured students' spatial reasoning to ensure that differences in performance gains are not driven solely by differences in students' spatial skills. The pre- and posttest assessments are 22-item free response measures that probe

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

student understanding of stereochemical properties of organic molecules (see Supporting Information for the pre- and posttest instruments) and ask students to answer questions relating to these learning outcomes:

(1) identify the relationship between a pair of molecules (identical molecules, constitutional isomers, or stereoisomers);
(2) quantify the number of chirality centers in a given molecule;
(3) rank substituents as if they were attached to a chirality center;
(4) assign *R*/*S* configuration to a chirality center depicted with well-defined stereochemistry;
(5) draw a molecule in *R*/*S* configuration when stereochemistry was not defined;
(6) identify the relationship between a pair of molecules (enantiomers, diastereomers, or meso compounds); and
(7) draw an enantiomer or diastereomer of a molecule depicted with well-defined stereochemistry.

It is noted that the pre- and posttest instruments focus on skills that are zero- or one-dimensional, as defined by the three-dimensional learning assessment protocol.[16] These levels of questions were chosen because the limited time frame of the interventions might have led to more limited improvement in higher-dimension scientific and engineering practices and were chosen to reduce the subjectivity of evaluating the student performance on the pre- and posttests.

The flipped treatment and traditional lecture treatment groups completed the first learning interventions (day 4). For the flipped treatment group, students were given 90 min to individually watch a playlist of five videos (total video time: 23 min, 6 s) which were integrated with quiz content using the Playposit platform.[17] The Playposit questions were embedded within the video with the intent of keeping students engaged while content was presented. The videos were produced in a screen-capture system with instructor voiceover, featuring the same instructor as the traditional lecture treatment. These questions were presented in multiple-choice format, and students received feedback on incorrect answers. Students had control over playback (i.e., pause and review segments) and could watch the videos multiple times. Students were not able to ask clarifying questions to instructors nor to have discussion with each other. For the traditional lecture treatment group, students attended a 90 min in-class lecture given by the instructor. As the lecture was given by the same instructor that made the videos, the material featured was at the same level in each setting. Example problems were carried out by the instructor in the lecture treatment group, and these corresponded to the questions asked in the Playposit system.

The flipped and traditional lecture treatment groups then completed posttest 1 prior to the second learning intervention (day 5). During this 90 min period, students were given 20 min to complete the same pretest (posttest 1) they had taken on day 0. Subsequently, students in the flipped treatment were instructed to form small groups (2−4 students) and work collaboratively on a problem set. This problem set featured free response questions that addressed the same types of questions asked on the pretest/posttest assessments. During this time students could discuss problems among themselves and ask clarifying questions from the instructor who circulated around the classroom. Students in the traditional lecture treatment were instructed to work on the same problem set individually. To emulate traditional homework, students were not allowed to discuss problems with each other or the instructor but could check their work using a key that was provided. Finally, all study participants completed the final posttest (posttest 2) (day 7). Study participants did not have access to the pretest assessment given at the beginning of the study; thus, the pretest was administered again on the fifth day of the study to measure performance gains after the first day of treatment. This is denoted as "posttest 1". The final posttest contained different items than the pretest, but each item was analogous to the items on the pretest. Tests were scored by author M.D.C. using an objective answer key (see pre- and posttests in Supporting Information).

Statistical Analyses

In order to test the experimental hypothesis, student performances on the pre- and posttest assessments and spatial reasoning test were compared both within and between each study group over the three testing points. The SPSS (Statistics Package for the Social Sciences) Statistics 24 software package was used to carry out the one-way ANOVA and test validity statistics.[18] The *lmerTest* and *emmeans* packages in R version 3.5.3 were used to run a mixed-effects ANOVA model, which included the treatment groups as between-subjects factors and time as the within-subjects factor. The equation used for the mixed-effects model is as follows:

$$Y = \mathrm{B}_0 + B_1X_1 + B_2X_2 + B_3X_1X_2 + \varepsilon$$

where Y represents test score, X_1 represents condition, X_2 represents time (i.e., the three testing sessions), and ε represents the error. Tukey honestly significant difference (HSD) posthoc pairwise comparisons were also calculated in R to compare the performance within and between the study groups. Hedges's g effect sizes were calculated using group means and standard deviations on each pre- and postassessment for the within-group pairwise comparisons, and the between-groups effect sizes (d_{ppc2}) were calculated as described by Morris.[19] Hedge's g and d_{ppc2} effect sizes were interpreted as described by Cohen (effect sizes greater than 0.20 = small effect, greater than 0.50 = medium effect, and greater than 0.80 = large effect).[13]

Interview Methodology and Analyses

Two weeks following the study, five participants from each treatment group were randomly selected to participate in a follow-up interview. These interviews were conducted to obtain qualitative data that could be used to complement the quantitative data described above, and more specifically to gain further insight about which aspects of the treatment interventions students appeared to value most and learn how the treatments carried out in the study compared to a real-world classroom implementation. Interviews were audio recorded and then transcribed prior to analysis. Interview questions for the students in the flipped treatment condition and traditional lecture condition have been included in the Supporting Information (Supplemental Tables 7 and 8), respectively.

■ RESULTS

The pre- and posttests were administered to the study groups as described above, and the mean scores at each testing point are summarized along with the descriptive statistics of the study populations in Table 2. To reduce the possibility students might experience improved performance due to

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 2. Descriptive Statistics for Treatment and Negative Control Groups

Statistic	Negative Control	Flipped Treatment	Traditional Lecture Treatment
Number of participants (n)	n = 16	n = 15	n = 18
Pretest average[a]	7.12 ± 2.87	6.80 ± 2.93	7.00 ± 2.68
Posttest 1 average[a]	n/a	13.13 ± 3.68	12.22 ± 5.31
Posttest 2 average[a]	6.62 ± 2.68	13.33 ± 2.69	13.61 ± 3.35
MRT average[b]	8.68 ± 2.52	7.47 ± 3.80	9.11 ± 5.82
UCR GPA[c]	3.178 ± 0.478	3.023 ± 0.951	3.141 ± 0.911

[a]Pre- and posttest average scores are reported out of a total of 22 possible points. [b]MRT average scores are reported out of a total of 24 possible points. [c]Average GPA for all sophomores and juniors in the UCR College of Natural and Agricultural Science = 2.99 ± 0.52; n = 2,575 (GPA scale reported out of a total possible of 4.0).

familiarity with the test instrument, the specific chemical structures used in the posttest 1 question items were changed to create an analogous assessment that was then used as posttest 2. Because customized assessments had to be created to test the distinct set of learning objectives covered in the treatment group interventions, item-analyses were carried out and the internal assessment reliability was determined for both assessments by measuring coefficient α (see SI, Supplemental Tables 1–3). The items in both posttests were deemed to be within an acceptable range of difficulty (all but one of the item means ranged from 0.20 to 0.80 for both posttests; see SI, Supplemental Table 3), and all but one of the questions had positive item-total correlations above 0.20 for both posttests (see SI, Supplemental Table 2). Item 14 was found to have a negative item-total correlation in both tests; therefore, this question was removed prior to all analyses in which the pre- and posttest scores were compared within or between study groups. Because there were multiple content items included in the posttest instrument, the overall internal reliability of the posttest was assessed using the stratified α reliability coefficient (α_s) as described by Widhiarso and Ravand.[20] The internal reliability was found to be in the acceptable range for both posttests (α_s = 0.803 and 0.832 for posttest 1 and posttest 2, respectively; see SI, Supplemental Table 1). Though posttest 2 was not identical to posttest 1, the similarity in internal validities and the plots of item discrimination versus item difficulty (see SI, Supplemental Figure 1) provide evidence the two posttests were adequately aligned and similarly measured the student learning objectives covered in the study treatments.

Even though the random assignment of study participants should have minimized selection validity threats, a one-way ANOVA was carried out to determine if the two treatment populations had significant differences in incoming spatial reasoning skills. This analysis appears to suggest the null hypothesis stating the two populations had equivalent incoming spatial reasoning skills cannot be rejected (see SI, Supplemental Table 4); therefore, this independent variable was not controlled in the subsequent analyses.

Though the descriptive statistics suggest both treatment groups appeared to make most of their overall performance gains after the first day of the intervention (see Table 2), it was desired to determine the significance of the difference in mean test scores between each treatment group, and within each treatment group at the different time points. To compare the performance gains between all three conditions over the course of the intervention, a mixed-effects ANOVA model was run as described above. These results indicate a significant effect of condition, a significant effect of time, and a significant condition by time interaction (see SI, Supplemental Table 5.) Tukey HSD adjusted posthoc pairwise comparisons reveal that both the flipped treatment and the lecture conditions showed greater improvement in the assessment scores than the negative control group, but there was no statistically significant difference between the flipped treatment group and the lecture group in overall improvement (see Table 3). Furthermore,

Table 3. Tukey Adjusted Posthoc Pairwise Contrasts Examining Differences between Conditions[a]

Pairwise Comparison	Difference Estimate	df	t	p	d_{ppc2}[b]
Negative Control–Flipped	−4.297	46	−4.376	<0.001	2.338
Negative Control–Lecture	−4.153	46	−4.423	<0.001	2.477
Flipped–Lecture	0.144	46	0.151	>0.05	−0.029

[a]df = degrees of freedom. t = test statistic. p = significance. [b]Effect sizes calculated here are d_{ppc2} as reported by Morris for repeated measures designs; see ref 19.

both the flipped treatment group and the lecture group significantly improved from pretest to posttest 1 (see Table 4)

Table 4. Tukey Adjusted Posthoc Pairwise Contrasts Examining Differences within Conditions over Time[a]

Pairwise Comparison	Difference Estimate	df	t	p	g[b]
Negative Control: Pretest–Posttest 2	0.500	92	0.600	>0.050	0.176
Flipped Pretest–Posttest 1	−6.333	92	−7.354	<0.001	1.799
Flipped Posttest 1–Posttest 2	−0.200	92	−0.232	>0.050	0.088
Flipped Pretest–Posttest 2	−6.533	92	−7.587	<0.001	2.195
Traditional Lecture Pretest–Posttest 1	−5.222	92	−6.643	<0.001	1.186
Traditional Lecture Posttest 1–Posttest 2	−1.389	92	−1.767	>0.050	0.336
Traditional Lecture Pretest–Posttest 2	−6.611	92.0	−8.410	<0.001	2.084

[a]Hedges's g effect sizes are reported for each pairwise comparison (df = degrees of freedom; t = test statistic; p = significance). [b]Hedge's g is similar to Cohen's d, except it is calculated using sample-based pooled standard deviations rather than population-based pooled standard decisions.

but did not show statistically significant improvement from posttest 1 to posttest 2 (see Table 4). As expected, the negative control group showed no change in performance over the course of the study. Finally, there were significant performance gains for the flipped treatment condition and lecture condition from pretest to posttest 2 (see Table 4). A graph illustrating student performance between the three conditions and across the three testing time points is shown in Figure 2.

Interview transcripts were coded for the following themes: (1) student motivation/reasoning for why they participated in the study; (2) which learning component was perceived as most useful (e.g., online learning, in-person practice prob-

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

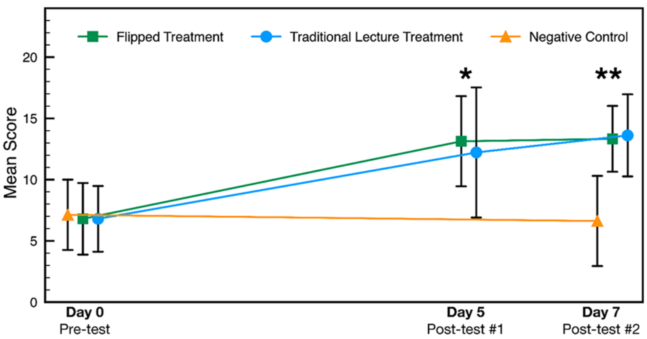

Figure 2. Mean performance (maximum score: 22) on the measure of stereochemistry understanding at all three testing time points (pretest, posttest 1, and posttest 2) for each condition (lecture, flipped treatment, and negative control). *Indicates posttest 1 scores for the flipped and traditional lecture treatment groups were significantly higher than their corresponding pretest scores. **Indicates the posttest 2 scores for the flipped and traditional lecture treatment groups were significantly higher than the posttest 2 score for the negative control group.

lems); and (3) the challenges students faced in their assigned learning condition. All interview data were dual coded by author K.A. and by a research assistant. After coding the transcripts for the themes, the coders summarized each of the participants' responses (provided in SI, Supplemental Tables 7 and 8). There was 90% agreement between the two coders, but areas of disagreement in the coding were resolved through discussion. We highlight here the following results: (1) four out of five participants from each treatment group reported they were motivated to participate in the study because of the opportunity to gain extra practice in learning stereochemistry; (2) four out of five flipped participants stated their favorite part of the study was using the online videos; (3) three of the five flipped participants stated they preferred the flipped classroom format to a traditional lecture; (4) four out the five flipped participants stated they would have liked to spend more time watching the videos; and (5) three out of the five flipped participants felt the collaborative group learning was fundamental to their learning the stereochemistry concepts (see SI, Supplemental Tables 7 and 8).

■ DISCUSSION

This study was designed to gain insight into which aspects of the flipped classroom might have the most impact on student performance. The topic of stereochemistry in organic chemistry was specifically chosen because it is a concept that students do not encounter prior to the organic chemistry course, and all students would start with minimal pre-existing knowledge related to these learning objectives. The general effectiveness of the treatment groups is evident from the descriptive statistics, which reveal notable increases in the posttest scores relative to the pretest (see Table 2, Table 4, and Figure 2). It is particularly noteworthy that even after a single day of learning, significant improvement in the posttest scores was observed in both treatment groups. This is evidenced by the pretest to posttest 1 gains relative to the negative control. However, the flipped classroom treatment appears to have made no additional gains after the second day of learning, whereas the traditional lecture appears to have made some additional minor gains from posttest 1 to posttest 2 (see Table 2, Table 4, and Figure 2).

Experimental Hypothesis and Research Question 1: Online Learning versus Group Learning

The statistical significance of the differences in mean test scores between both groups, and within groups over time, are revealed in the mixed-effects ANOVA model and posthoc tests described above. These results clearly indicate significant performance gains were made after the first day of learning for both treatment groups, as the within-group pairwise comparisons indicate that posttest 1 scores were significantly higher than the pretest scores (see Table 4 and Figure 2). Conversely, the within-group pairwise comparisons suggest negligible performance gains were made after the flipped treatment collaborative group learning or traditional lecture treatment individual homework problem solving (see Table 4 and Figure 2). Therefore, it appears that under these experimental conditions the experimental hypothesis cannot be rejected, in which it was stated the preclass online learning component of a flipped classroom contributes as much, or more than, the in-person collaborative group learning to the overall performance gains in a model flipped classroom intervention. Interestingly, the passive lecture also appeared to account for most of the improvement in posttest scores in the traditional lecture treatment; however, the effect sizes suggest the online video treatment may have led to greater performance gains relative to the passive lecture (g = 1.799 and 1.186, respectively; see Table 4). It is also noted that though the differences in mean scores for posttest 1 and posttest 2 were not statistically significant, the effect sizes suggest the individual homework intervention may have had a greater subsequent impact on student performance than the collaborative group learning (g = 0.336 and 0.088, respectively, see Table 4).

Classroom studies that compare flipped and traditional approaches to teaching often compare exam scores between groups, but to our knowledge the report from Moore and coworkers is the only study in which the impact of the different aspects of a typical flipped classroom were measured in parallel.[12] Though this previous study found a fully flipped classroom appeared to yield higher exam scores compared to a nonflipped class that included collaborative group learning, the performance gains that took place within the fully flipped classroom were not correlated to any particular part of the flipped classroom intervention. The structure used in the present study focused on teaching a learning objective over an abbreviated time period outside a for-credit classroom environment, which provided an opportunity to assess which aspects of the treatments resulted in the greatest improvement in posttest scores. The results here appear to corroborate the general findings from Moore, in which it was found the preclass learning component of the flipped classroom intervention is an important contributor to student performance. Sweller's cognitive load theory and Mayer's cognitive load theory of multimedia learning suggest preclass multimedia interventions should reduce cognitive load relative to a traditional lecture. This is primarily achieved by integrating and synchronizing auditory and visual components in the video presentation (versus students watching an instructor from afar, and successively receiving auditory and visual input that is then transcribed into notes). The online learning environment used in this study also reduced cognitive load by segmenting the presentation of content, both by taking the 90 min lecture and breaking it down into multiple shorter videos, but also by segmenting the student work within each video by integrating the Playposit questions. Cognitive load theory also suggests

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

students learn to "chunk" together different elements of information that can be processed as one entity in their working memory, and that these chunks reside in the long-term memory or are formed using knowledge from long-term memory.[5,7] It is also thought that novices have less developed chunks of knowledge in their long-term memory;[5,7] therefore, it is quite possible the segmented video tutorials typical in the flipped classroom online learning environment can help students overcome this cognitive limitation.

Though the difference in posttest 1 mean scores between the flipped classroom and traditional lecture treatments did not appear to be statistically significant, the flipped classroom treatment did have a higher mean score on posttest 1 with a higher effect size (flipped treatment, g = 1.799; traditional lecture treatment, g = 1.186; see Table 4). This suggests the online learning intervention may have indeed improved student performance relative to the passive lecture, but future research needs to focus on designing online learning interventions that can further reduce cognitive load for students in an effort to maximize the impact of this learning environment. We also note that four of the five students interviewed from the flipped treatment group indicated they would have spent more time engaging in the online video environment; however, this was not possible due to the limitations of the experimental design. This implies cognitive load could have been further reduced had the flipped classroom intervention more closely modeled a real classroom implementation. Finally, it is noted the students who volunteered for this study appeared to have higher GPAs relative to the UCR College of Natural and Agricultural Science overall student population (see Table 2). Since flipped classroom interventions have been found to have the most impact on less prepared students,[9,10] the fact the students in this study appear to be higher achieving students may have also limited the impact of the flipped classroom treatment relative to the traditional lecture treatment.

Research Question 2: Flipped Sequence versus Traditional Sequence

On the surface, it was somewhat surprising that the flipped classroom treatment appeared to make no additional performance gains after the collaborative group learning intervention, but a more careful analysis framing this experimental implementation within Hogan's theory of knowledge coconstruction might explain these results.[6] This theoretical framework suggests that improved performance is linked to how students form relationships with their peers and instructor. Because the collaborative group learning conducted in this study was restricted to a single 1.5 h long intervention, longer-term social relationships could not be built within student groups or with the instructor. This might have stunted the performance gains that would normally be expected to arise from this type of learning environment.[21,22] Another possible explanation for the plateauing of student posttest scores in the flipped treatment may be linked to the assessment itself. Even though students had room to make additional performance gains (the average score on posttest 2 was 13.33 out of 22 possible points), perhaps the remaining question items were simply too difficult for the students to learn within the time frame of the study. With that said, students who completed individual homework in the traditional lecture treatment did appear to make slightly larger performance gains from posttest 1 to posttest 2 (see Table 2 and Table 4), and item-analyses of posttest 1 and posttest 2 within the flipped treatment group indicate there was some variability in the distribution of missed test items among the study participants (e.g., item 3C had a mean = 0.0882 in posttest 1, whereas it had a mean = 0.206 in posttest 2; and item 6B had a mean = 0.500 in posttest 1, whereas it has a mean = 0.794 in posttest 2; see SI, Supplemental Table 3). These data suggest the lack of student performance gains was most likely attributed to a limitation in the design of the intervention and not attributed to the assessment used in measuring student understanding of stereochemistry.

Limitations and Future Work

Future research will aim to address some of the limitations noted for this study. In particular, longer study times will be planned in an effort to better model how students use the online learning space, and to increase the potential impact of the collaborative group learning environment. Additionally, studies will be designed to measure how online learning environments might be optimized for different types of learning outcomes (e.g., spatial reasoning versus algorithmic problem solving) and how the structure of the collaborative group learning impacts this part of the flipped classroom intervention. These follow-up studies will explore different implementations of online learning environments to probe which are most effective for each type of learning outcome. Student comments from the focus group interviews suggest the flipped treatment participants found stereochemistry to be a fairly easy topic, while the traditional lecture described it as more intermediate in nature. This may point to the online learning environment being of greater effectiveness for spatial learning outcomes. A recent study by Stull and Hegarty provides evidence that interventions using hand-held models or computer-based virtual models were superior to teaching structural representations for organic molecules relative to an intervention in which students received verbal feedback on paper-based drawing assignments.[23] It was also found that the handheld models and computer-based models reduced cognitive load and helped students scaffold their knowledge, which led to more sustained learning relative to the paper-based drawing intervention. Though the report from Stull and Hegarty did not explore the use of molecular models in an online environment, these results suggest online modules that incorporate computer-based interventions can act as an effective platform for teaching molecular structure and submicroscopic representations.

In addition to probing how to most effectively design online learning modules for different areas of chemistry content, it will be prudent to carry out experiments in which higher-dimensional thinking is incorporated into the online learning interventions. It was noted earlier that the learning objectives included in this study were determined to be zero- and one-dimensional skills within the three-dimensional learning assessment protocol (i.e., the learning objectives chosen for the treatment groups were foundational skills needed to perform future higher-dimensional tasks).[16] To our knowledge studies that compare the effectiveness of online learning platforms in teaching and assessing concepts at multiple levels of the three-dimensional learning protocol have not been reported. Therefore, future studies will also include online learning modules that incorporate scientific and engineering practices, cross-cutting concepts, and/or disciplinary core ideas. Demonstrating that flipped classrooms can help students

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX

achieve learning that goes beyond rote learning and basic skill building will be critical in providing proof-of-concept for chemistry practitioners interested in adopting the flipped classroom structure.

Conclusion

In summary, the findings described herein suggest historical implementations of the flipped classroom may have overlooked the impact of the asynchronous online learning environment on student performance outcomes relative to in-person collaborative group learning. In terms of the implications for classroom practitioners, the results presented herein provide evidence that students can indeed make significant gains in skill-based assessments in an asynchronous online learning environment. Perhaps these results might alleviate fears some classroom instructors have in releasing their students into this more individualized and unsupervised learning space. Though this study was not designed to compare the efficacy of a model flipped classroom to a traditional lecture intervention, previous quasi-experiments have demonstrated that flipped classrooms generally lead to improved performance on summative assessments and higher course grades, and chemistry instructors should strongly consider adopting this classroom structure in lieu of more passive lecture pedagogies. More importantly, the findings presented herein suggest the chemistry education community needs to focus its attention on improving the design of the online learning environment within the flipped classroom scaffold. If educational researchers and instructional designers can focus more attention on optimizing student learning for different types of learning objectives within an asynchronous online space, significant progress on improving student learning outcomes and success in introductory undergraduate STEM courses might be achieved. Given the rapidly expanding use of the flipped classroom approach, making this type of targeted improvement in student learning within the online learning space is likely to positively impact the broader success and retention of students in STEM gateway courses.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.9b00767.

Supplemental data tables and figures (PDF, DOCX)

Pretest and posttest 1 instrument and key (PDF)

Posttest 2 instrument and key (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: matthew.casselman@ucr.edu.
*E-mail: jack.eichler@ucr.edu.

ORCID

Matthew D. Casselman: 0000-0001-9040-4663
Jack F. Eichler: 0000-0003-2755-3636

Author Contributions

§M.D.C., K.A., and J.F.E. contributed equally to this work.

Notes

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research is based upon work supported by the National Science Foundation under Grant 1504989.

■ REFERENCES

(1) Seery, M. K. Flipped learning in higher education chemistry: Emerging trends and potential directions. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 758−768.
(2) Deslauriers, L.; Schelew, E.; Wieman, C. Improved learning in a large-enrollment physics class. *Science* **2011**, *332* (6031), 862−864.
(3) Crouch, C. H.; Mazur, E. Peer Instruction: Ten years of experience and results. *Am. J. Phys.* **2001**, *69* (9), 970−977.
(4) Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; Wenderoth, M. P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (23), 8410−8415.
(5) Sweller, J. Cognitive load during problem solving: Effects on learning. *Cognitive Science.* **1988**, *12* (2), 257.
(6) Hogan, K. Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *J. Res. Sci. Teach.* **1999**, *36* (10), 1085−1109.
(7) Mayer, R. E.; Moreno, R. A cognitive theory of multimedia learning: Implications for design principles. *Journal of Educational Psychology.* **1999**, *91* (2), 358−368.
(8) Greer, D. L.; Crutchfield, S. A.; Woods, K. L. Cognitive theory of multimedia learning, instructional design principles, and students with learning disabilities in computer-based and online learning environments. *Journal of Education.* **2013**, *193* (2), 41−50.
(9) Ryan, M. D.; Reid, S. A. Impact of the Flipped Classroom on Student Performance and Retention: A Parallel Controlled Study in General Chemistry. *J. Chem. Educ.* **2016**, *93* (1), 13−23.
(10) Crimmins, M. T.; Midkiff, B. High Structure Active Learning Pedagogy for the Teaching of Organic Chemistry: Assessing the Impact on Academic Outcomes. *J. Chem. Educ.* **2017**, *94* (4), 429−438.
(11) Eichler, J. F.; Peeples, J. Flipped classroom modules for large enrollment general chemistry courses: a low barrier approach to increase active learning and improve student grades. *Chem. Educ. Res. Pract.* **2016**, *17* (1), 197−208.
(12) Rau, M. A.; Kennedy, K.; Oxtoby, L.; Bollom, M.; Moore, J. W. Unpacking "Active Learning": A Combination of Flipped Classroom and Collaboration Support Is More Effective but Collaboration Support Alone Is Not. *J. Chem. Educ.* **2017**, *94* (10), 1406−1414.
(13) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: New York, NY, 1988; Chapter 9.
(14) Vandenberg, S. G.; Kuse, A. R. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills.* **1978**, *47* (2), 599−604.
(15) Harle, M.; Towns, M. A Review of Spatial Ability Literature, Its Connection to Chemistry, and Implications for Instruction. *J. Chem. Educ.* **2011**, *88*, 351−360.
(16) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11*, e0162333.
(17) Playposit. https://go.playposit.com/ (accessed October 21, 2019). Playposit is a web-based platform that allows instructors to embed lecture videos within an assessment system. This provides an opportunity to make the online learning more engaging for the students, allows instructors to reward class credit for completion of the videos, and gives students more feedback while completing these asynchronous activities.

(18) IBM Corp. *IBM SPSS Statistics for Windows*, Version 24.0; IBM Corp: Armonk, NY, 2016.
(19) Morris, S. B. Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods.* **2008**, *11* (2), 364−386.
(20) Widhiarso, W.; Ravand, H. Estimating reliability coefficient for multidimensional measures: A pedagogical illustration. *Review of Psychology* **2014**, *21* (2), 111−121.
(21) Chapman, C.; Ramondt, L.; Smiley, G. Strong community, deep learning: Exploring the link. *Innovations in Education and Teaching International* **2005**, *42* (3), 217−230.
(22) Brindley, J.; Blaschke, L. M.; Walti, C. Creating Effective Collaborative Learning Groups in an Online Environment. *International Review of Research in Open and Distributed Learning* **2009**, *10* (3), DOI: 10.19173/irrodl.v10i3.675.
(23) Stull, A. T.; Hegarty, M. Model Manipulation and Learning: Fostering Representational Competence with Virtual and Concrete Models. *Journal of Educational Psychology* **2016**, *108* (4), 509−527.

DOI: 10.1021/acs.jchemed.9b00767
J. Chem. Educ. XXXX, XXX, XXX−XXX