

Portland State University

PDXScholar

Chemistry Faculty Publications and
Presentations

Chemistry

8-14-2020

Multicourse Comparative Study of the Core Aspects for Flipped Learning: Investigating In-Class Structure and Student Use of Video Resources

Nicole Naibert

Portland State University

Emryse Geye

Portland State University

Michael M. Phillips

University of Northern Colorado

Jack Barbera

Portland State University, jbarbera@pdx.edu

Follow this and additional works at: https://pdxscholar.library.pdx.edu/chem_fac



Part of the [Chemistry Commons](#), and the [Science and Mathematics Education Commons](#)

Let us know how access to this document benefits you.

Citation Details

Published as Naibert, N., Geye, E., Phillips, M. M., & Barbera, J. (2020). Multicourse Comparative Study of the Core Aspects for Flipped Learning: Investigating In-Class Structure and Student Use of Video Resources. *Journal of Chemical Education*, 97(10), 3490-3505.

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Chemistry Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

A Multi-course Comparative Study of the Core Aspects for Flipped Learning: Investigating In-Class Structure and Student Use of Video Resources

^aNicole Naibert, ^aEmryse Geye, ^bMichael M. Phillips, and ^{a*}Jack Barbera

^aPortland State University – Department of Chemistry, Portland Oregon

^bUniversity of Northern Colorado – School of Psychological Science, Greeley Colorado

*jack.barbera@pdx.edu

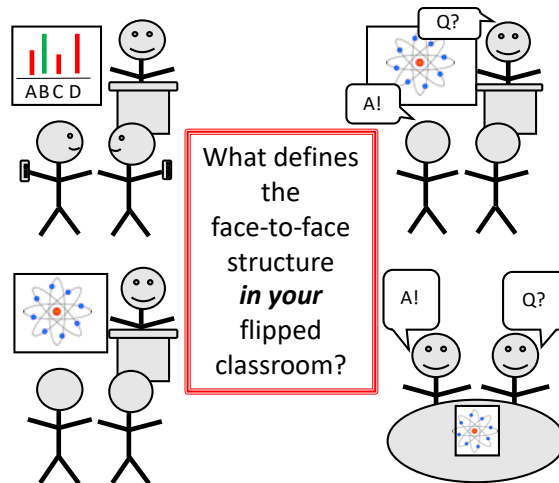
Abstract

Since 2013, the number of publications on flipped learning within chemistry have steadily increased. However, most of these studies focus on flipped course reforms within individual institutions, while the outcomes of any learning environment are dependent on how the environment is structured and the degree to which students interact with its elements. In this study, we apply a coordinated set of assessment practices to investigate similarities among flipped chemistry courses at five institutions in the United States. All courses in the study followed the two basic tenets of flipped learning: 1) foundational information was delivered through pre-class materials (PCMs) and 2) the face-to-face (F2F) environment applied or extended the content through active learning. Each F2F environment was characterized using video recordings analyzed with the Classroom Observation Protocol in Undergraduate STEM (COPUS) tool. Each individual course showed consistent use of F2F time across each session recording, however, there were significant differences in the predominant student behaviors between courses. Student behavior in two of the courses (Courses Four and Five) was dominated by work in small-groups on problem solving worksheets, in contrast to another course (Course One) where responding to whole-class questioning posed by the instructor dominated students' behaviors. While behaviors in the two remaining courses (Courses Two and Three) included a mix of responding to clicker and whole-class questions, one of them (Course Three) also included large episodes of students simply listening during instructor presentation of material. A mid-semester survey was administered in each course to characterize students' interactions with, and perceptions of, the PCMs. Of particular note, student self-reports of the number of videos viewed and the timing of viewing trended with the amount of peer-to-peer interaction during F2F sessions. That is, students in courses with more consistent groupwork reported watching more of the video content and doing so before the F2F session. These results demonstrate that flipped classrooms can take many forms and suggest that F2F structure may create non-grade-based incentives for PCM utilization.

Reference Information:

Nicole Naibert, Emryse Geye, Michael M. Phillips, and Jack Barbera, "Multi-course Comparative Study of the Core Aspects for Flipped Learning: Investigating In-Class Structure and Student Use of Video Resources," *Journal of Chemical Education*, 2020, 97, 3490-3505, DOI: 10.1021/acs.jchemed.0c00399.

Graphical Abstract



Keywords: First-Year Undergraduate / General, Chemical Education Research, Collaborative / Cooperative Learning, Student-Centered Learning

Introduction

A national call from the President's Council in 2012¹ to increase the number of science, technology, engineering, and mathematics (STEM) degrees encouraged instructors to take a closer look at how higher-education STEM classrooms supported student learning. A subsequent report from the National Research Council on Discipline-Based Education Research (DBER)² and the Freeman et. al.³ meta-analysis, further promoted the inclusion of more active learning activities in higher-education STEM classrooms. However, one barrier to including more active learning is the amount of time it takes to include these activities in the classroom.⁴ To address this barrier, some instructors “flip” their classroom, allowing significant flexibility in how to structure the course. Most flipped classrooms follow two basic tenets: 1) foundational information is delivered to students through pre-class materials (PCMs), and 2) the face-to-face (F2F) environment is used to apply or extend the information through active learning.^{5, 6} Therefore, a “flipped” classroom provides the instructor with flexibility in the type of PCMs provided to the students, as well as the type of in-class active learning conducted in the F2F environment.

Structure of flipped learning environments

Higher-education flipped chemistry classrooms include a variety of different F2F active learning environments, such as Peer-Led Team-Learning (PLTL),⁷⁻¹⁰ Process-Oriented Guided-Inquiry Learning (POGIL),^{11, 12} or other combinations of peer instruction, problem-based learning, and/or think-pair-share type exercises.^{13, 14} In addition, some classes incorporate case studies^{9, 15} or whole-class discussions¹⁶⁻²¹ into their class time. Most incorporate some type of groupwork²²⁻³⁴ and several use mini-lectures or Just-in-Time Teaching to provide clarification on concepts when needed.^{10, 13, 20, 22-25, 27, 30-33, 35} Some provide additional time for these activities by adopting a hybrid structure where students participate in active learning during a specific block of time or day of class.^{7-10, 18, 30}

Regardless of the type of F2F learning environment, flipped classes are structured such that students enter the classroom with some level of understanding, which they build upon during class time.^{6, 36, 37} Often the assumption is that this understanding comes from PCMs that students complete before class. Abeysekera and Dawson³⁸ noted that a benefit of this tenet of flipped learning is the ability for students to self-pace their learning and therefore manage cognitive load.³⁹ Stemming from this early acknowledgement of cognitive load as an underpinning framework to understand the benefits of flipped learning, cognitive load theory⁴⁰ has been used to provide direct support for the design of a flipped course.⁸ By allowing students the opportunity to use PCMs on their own time and at their own pace, it tends to reduce the information overload burden.⁴⁰ In addition, students have the potential to re-watch PCMs, if needed.⁴¹ The design of PCMs⁴¹ have also been supported through the cognitive theory of multimedia learning.⁴² PCMs in most flipped chemistry courses entail students use of some type of online component with dual aspects targeted to both auditory and visual information, with instructor-made videos (e.g., screencasts) being the most common.^{8-10, 12-14, 19-21, 23-25, 27, 31, 33, 34, 41, 43-46} However, other online resources, such as Khan Academy videos⁷ or interactive online modules²⁹ have also been provided as PCMs. Further, some flipped classes include multiple formats of PCMs, such as videos, screencasts, interactive modules, animations, podcasts, etc.^{11, 16, 18, 22, 28, 35, 47}

Student utilization of pre-class materials (PCMs)

Instructors may provide incentives to encourage students to complete PCMs on time; for example, checking that students have taken required notes on the PCMs.^{12, 31, 33} Many utilize quizzes administered either in-class^{13, 21, 24, 25, 28, 46} or out-of-class as part of the required PCMs.^{8, 10, 14, 18, 19, 22, 23, 25, 27, 34, 35, 47} The purpose and difficulty of quizzes varies across studies, with some descriptions emphasizing that quizzes are scored primarily for completion points and/or formative assessment^{18, 23, 24, 27, 34, 46} or that students are given an opportunity to discuss answers before submitting.²¹ Regardless of the type and level of incentive, PCMs are provided to students to acquire initial content understanding, which can be built upon during F2F activities. Thus, determining “if”, “how”, and “when” students are using the assigned PCMs may be important for characterizing the success of a flipped learning environment. However, not all studies on flipped chemistry courses include this type of information. Some have determined “if” students completed the PCMs by keeping track of who turned in problem-solving activities,³⁰ by asking students if they watched the required videos for that day,⁴⁶ or by assessing students’ performance on quiz questions based on non-conceptual topics embedded in the videos.²⁴ Some studies include self-reported student estimates of time spent studying outside of class each week²⁹ or specifically watching the videos.³³ To gain information about “how” and “when” students interact with PCMs, more detailed survey questions have been administered to students asking about the frequency and nature of video use and of the videos’ perceived usefulness, length, quality, and impact on the class.²¹ Details about “if”, “how”, and “when” students interact with PCMs has also been reported through the use of analytics data collected from student access tracking^{8, 23, 31, 33, 34, 43} or other video tracking data.^{8, 33, 43} In addition, analytics data has been used to provide information about “if” and “when” students were re-watching videos.^{8, 33}

Students’ use of PCMs can vary greatly from class to class. For example, Lenczewski³⁰ noted that “*about 33% of the class completes 80% or more of the assignments per semester*” and Bancroft et al.²³ reported that although about 98% of students watched the assigned videos

before class at the beginning of the term, this number dropped to around 85% by the end. Mooring et al.⁸ reported that 80-95% of students watched the assigned videos and completed the incentive quiz, with Woodward and Reid⁴⁶ noting that they only obtained 79% viewership without regular email reminders. Thus, students make use of the PCMs to varying degrees. Therefore, depending on the structure of the F2F environment, students may enter into learning activities with insufficient understanding to fully participate.

Purpose and Rational

This project is part of a larger study on flipped learning environments and their impact on student motivation and performance. The first phase of this coordinated, multi-institution study involves evaluating 1) the structure of each F2F learning environment and 2) students' use and perceptions of the PCMs that support these environments.

In a recent meta-analysis, Rahman and Lewis⁴⁸ reported on the effectiveness of a range of evidence-based instructional practices (EBIPs), including flipped learning. Their analysis of fifteen flipped learning studies revealed 'trivial to medium' effect sizes. In addition, they determined how the effect sizes for an EBIP were moderated by other factors, such as the type of assessment used and course size, and noted that an additional moderator for flipped learning may be variation in how the environment is structured. In acknowledgement that flipped structures vary,⁴⁹ Seery noted that in some studies "*there is a vagueness about what happens during class time, and a more robust framework needs to be developed so that there is a basis for what happens in class time and how it builds on pre-lecture work*". The call for more robust frameworks of EBIPs, in general, extends beyond Seery's specific call. In 2016, Stains and Vickrey⁵⁰ noted that the findings of EBIP studies are compromised if factors related to implementation cannot be accounted for. To address this, they proposed the use of a fidelity of implementation (FOI) framework. FOI studies have taken many forms within discipline-based education research (see examples in Stains and Vickrey⁵⁰), perhaps the most salient form for flipped learning is the investigation of how course structure impacts outcomes. While many flipped chemistry course studies provide descriptions of the F2F environment and the type of active learning activities employed during class time, few report observational data for consistent cataloguing of the structure. Of those that have, Canelas et. al.¹¹ categorized the percentage of classroom time that students spent participating in active learning and Donnelly and Hernández²⁶ used the Behavioral Engagement Related to Instruction (BERI) protocol to determine the percentage of students who were behaviorally engaged throughout class time. As there is no single prescribed structure for the F2F portion of a flipped classroom and the positive impacts of active learning (e.g., higher exam scores, lower failure rate, etc.) have been linked to students being more engaged with course activities,^{3, 51} it is important to understand the degree to which a given F2F structure is student-centered and the nuanced differences that exist. Therefore, we conducted consistent evaluations of each flipped learning environment such that the structure of their F2F settings can be accounted for when comparing course outcomes.

Recent studies on active learning have noted the association of student buy-in with self-regulated learning and course performance,⁵² and how students' perceptions may influence their buy-in to the learning environment.^{52, 53} Cavanagh⁵² notes that "*active learning provides students with opportunities to engage in the learning process, and students may decide to participate based on a series of judgements*", one of which is whether they view an activity as valuable to

their learning. These judgements have been shown to be influenced by the classroom climate,^{54, 55} and students may change their initial expectations within a course based on teaching practices and course demands.⁵⁶ Within flipped chemistry courses, lack of engagement with PCMs may undermine the potential benefits of the learning environment.^{6, 57} Incentivizing timely interaction with the PCMs have been noted to influence students' utilization of them.^{25, 46} In addition, the degree of a course's structure has been attributed to increased utilization of PCMs.⁵⁴ In defining "low-", "moderate-" and "high-structure" courses, one parameter used by Eddy et. al.⁵⁴ was "in-class engagement". This parameter was broken down by the amount of student-talk occurring during an F2F session through the use of activities such as clicker questions, worksheets, or case studies. By this parameter, a "low" course was defined as having student-talk for <15% of F2F time, "moderate" courses as 15-40%, and "high" courses as >40% of F2F time. Other parameters used when defining the degree of a course's structure included the frequency of the preparatory and review assignments. While their study did not include any high-structure courses, when comparing students' behaviors and perceptions between courses with low- and moderate-structure, Eddy et. al.⁵⁴ found that students in the moderate-structure course reported studying for more hours per week, completing recommended reading before class, and perceiving the preparatory work to be more important. The authors believe that the increased course structure led to more "accountability" on the students' part for their learning, thereby improving course performance in this as well as other studies.⁵⁸ These outcomes are in line with the ability for classroom environments to stimulate the development of self-regulated learning⁵⁹ and that self-regulated learners are empowered to create goals, use strategies, and implement actions to meet their goals.⁶⁰

Given the many ways in which a flipped learning chemistry course can be structured, and to move beyond a more general description of flipped learning as being defined as providing PCMs and more active learning during F2F sessions, this study employed a coordinated set of evaluation techniques (i.e., classroom observations and a survey of students' use and perceptions of the PCMs) across multiple flipped chemistry courses to answer the following research questions:

- 1) What are the predominant student behaviors and instructional styles for the F2F settings within these environments?
- 2) What are the students' self-reported use and perceptions of the PCMs in these environments?
- 3) What associations exist between instructional style in the F2F setting and PCM utilization?

Methods

Course Descriptions

Flipped chemistry courses from five institutions across the United States were involved in this study, none of which were the authors' home institutions. Institutions varied by size, type, acceptance rate, and demographic profile (Table 1). Four of the institutions were four-year public research universities and the fifth was a two-year community college. These data collection sites were selected based on the corresponding author's knowledge of who the instructors were and that they were not new to course flipping. Each invited instructor had a minimum of two years of experience in flipping their course and was the primary person involved in developing the course materials. The general structure of each course followed the two basic tenets of flipping: 1) foundational information was delivered to students through pre-class materials (PCMs), and 2)

the face-to-face (F2F) environment was utilized for the application or expansion of the information through active learning.^{5,6}

Both General and Introductory Chemistry courses were included in the study. Multiple sections of each course, taught by the same instructor, were combined in the datasets (Table 1). Course schedules and settings varied, the most notable of which is Course Five, which took place only once per week (for 75 min) in a fully collaborative space. The remainder of the courses were held on more traditional 2- or 3-day per-week schedules in either standard fixed seating lecture halls or ones designed with rotating chairs to promote collaborative work.

Table 1. Course and institution details.

Course	One	Two	Three	Four	Five
Type	General I	Introductory I	General II	General I	General I
Cycle	On sequence	On sequence	On sequence	Off sequence	On sequence
Sections	1	2	1	2	2
^a Enrollment	200	72	281	360	171
Schedule	75 min, 3 times per week, morning	50 min, 3 times per week, morning	80 min, 2 times per week, evening	80 min, 2 times per week, morning	75 min, once per week, afternoon
Setting	Auditorium style – rotating chairs	Auditorium style – fixed chairs	Auditorium style – rotating chairs	Auditorium style – rotating chairs	Collaborative space – circular tables of 9
Institution					
Region	Southeast	Southwest	Southwest	Northwest	Midwest
Size (Approx.)	55,000	10,000	35,000	30,000	50,000
Type	Four-year, Public, Doctoral – Very High Research Activity	Two-year, Public, Associate’s College – Mixed Transfer/Career & Technical	Four-year, Public, Doctoral – Very High Research Activity	Four-year, Public, Doctoral – Very High Research Activity	Four-year, Public, Doctoral – Very High Research Activity
Acceptance	50%	100%	30%	78%	45%
^b Demographics	Asian – 4% Black – 13% Hispanic – 60% White – 13% Other – 7%	Asian – 9% Black – 5% Hispanic – 25% White – 54% Other – 7%	Asian – 27% Black – 4% Hispanic – 12% White – 39% Other – 18%	Asian – 7% Black – 1% Hispanic – 9% White – 61% Other – 22%	Asian – 9% Black – 4% Hispanic – 3% White – 65% Other – 19%

^aTotal enrollment across all sections at the start of the term. ^b‘Other’ category can include designations of International, Pacific Islander, 2+ ethnicities, and/or other designations inconsistently reported across institutions.

Pre-Class Material (PCM) Description

Within each course, instructors assigned videos that corresponded to the content of each F2F class day. With the exception of Course One, in which the instructor curated relevant online videos from different sources, courses used instructor-created content videos (Table 2). Each instructor noted that their number and length of videos varied by topic across each term. Students completed an online quiz that covered the related video content prior to each F2F day in Courses One, Three, and Four. In Course Two, each F2F day started with a clicker quiz based on the video content. No grade-based viewing incentive was utilized in Course Five.

Table 2. Topics observed and video content details for each participating course.

Course	Topics observed	Video type	^a Number of videos assigned	Length range (minutes)	Majority length range (minutes)	Viewing Incentive
One	Periodic trends	Instructor curated online videos	2	6 – 11	6 – 11	Online quiz
Two	Oxidation numbers, activity series, and types of reactions	Instructor created screencasts	11	3 – 12	5 – 10	Clicker quiz
Three	Intermolecular forces, phase diagrams, and solid structures	Instructor created screencasts	10	2 – 17	5 – 10	Online quiz
Four	Lewis structures and formal charge	Instructor-created screencasts and recorded chalk-talks	4	11 – 25	15 – 20	Online quiz
Five	Electron configuration, periodic trends, and bonding	Instructor-created screencasts and recorded chalk-talks	9	15 – 44	20 – 25	None

^aEach course was observed on two consecutive class periods, these numbers correspond to the total across the observation days.

Within each participating course, two modes of data collection were employed: 1) observations of the F2F learning environments, and 2) online surveys administered to enrolled students. As the two data sources have different data collection protocols, the remainder of the methods section, as well as the results section, will be organized by those two sources. All data collected within this study was approved by the Institutional Review Board (IRB) at Portland State University and appropriate consent was acquired from instructors and students as required by the IRB.

Part 1: Structure of Face-to-Face (F2F) Learning Environment

Course Observations

Within each course, F2F sessions were recorded by a member of the research team on two consecutive class meetings. Recordings took place midway through a term on a non-exam week. Recordings were captured from the back of the room in a location where a large swath of the students could be seen as well as any primary location for the instructor (e.g., lectern or board). Throughout a recording, the researcher could zoom in-out and pan the camera to capture both instructor and student behavior as needed. The associated audio captured the majority of the instructor's talk, with the exception of their close-contact conversations with individuals or groups. Additionally, students' whole-class questions were captured but close-contact group conversations were not. Due to the logistics of planning on-site data collection, instructors were aware of which days they would be observed. With the variety of courses and the timing of site-visits, we were unable to observe the same content coverage across courses.

Coding Protocol

The Classroom Observation Protocol in Undergraduate STEM (COPUS⁶¹) was developed for use in higher education STEM, making it an appropriate choice for our needs. The protocol consists of codes used to document the real-time behaviors of both instructors and students during an F2F classroom session. Analysis of COPUS codes has been used to determine different instructional styles from lecturing and Socratic to peer instruction and collaborative learning⁶² and to establish instructional profiles called COPUS Profiles.⁶³ These profiles range from “didactic”, where more than 80% of the F2F time is used for lecturing, to “interactive lecture”, where the use of student-centered strategies emerges, and finally to “student-centered”, where the use of groupwork dominates the F2F time.⁶³

Each video was coded by two researchers using the COPUS. Each researcher was trained following the recommendations provided by the COPUS developers.⁶¹ After independently coding the first video for a given course, the coders met to compare their consistency in code use. This initial meeting provided clarity to the coders on how the various COPUS categories were defined relative to the course being observed. If observed behaviors in a course did not clearly align with code definitions (Table S1), the coders discussed and reached consensus on if or how the definition applied. Details on the code descriptions can be found in the Supporting Information. After these discussions, the coders recoded the first video before moving to the second. This coding process produced Cohen’s kappa scores >0.85 for each video, indicating high inter-rater reliability.⁶⁴

In addition to reviewing the individual COPUS categories, the data from each course were entered into the COPUS Analyzer.⁶⁵ This tool was developed during a large national study of STEM teaching practices⁶³ and provides COPUS Profiles for a given course based on a reduced set of observation data (4 instructor and 4 student codes). The COPUS Analyzer matches individual course data to one of seven clusters generated from a latent profile analysis conducted during the national study. Clusters are labeled as representing a various instructional style, either didactic lecture, interactive lecture, or student-centered. As described by the analyzer developers,⁶³ a didactic lecture style “*depicts classrooms in which 80% or more of class time consists of lecturing*”, whereas an interactive lecture style represents “*instructors who supplement lecture with more [compared to didactic instruction] student-centered strategies...such as clicker questions with groupwork*”. The student-centered style “*depicts instructors who incorporate student-centered strategies into large portions of their classes*”. In this study, clusters for a given course did not vary across observation days.

Part 2: Students’ Perceptions of Pre-Class Materials (PCMs)

Survey Development

The survey was developed during previous semesters through an iterative process that included two rounds of focus groups and one round of a pilot survey that included open-ended responses. This process informed the wording of items and response options. The final version of the survey was created such that each item contained multiple response options where students could choose a single option or multiple options, depending on the item type. Open response boxes were provided only if an ‘Other’ response was selected for an item. Brief descriptions of the development process are outlined in the Supporting Information.

Survey Administration

Survey participants were students recruited from the flipped courses to participate in the study. Survey deployment in each course was coordinated to take place midway through the term during a non-exam week. The instructor was provided a brief script to make an initial in-class announcement regarding the survey. A note similar to the script was posted on the classroom management platform of each course. Students who were interested in participating clicked on a link to the Qualtrics survey that was part of the announcement note. Some instructors offered a nominal amount of extra-credit points for accessing the survey.

Analysis

Data collected from the final version of the survey were cleaned to remove duplicate entries and any students who did not consent. For single-response items, the percentage of students selecting a given response was determined with responses to these items totaling to 100%, although there was some variability due to rounding. For multi-response items, the percentage of students choosing each response was determined. As students could select multiple options, responses are not mutually exclusive and may not total to 100%. As skip logic steps were built into the survey flow based on a student's response, not all students were presented each item. Therefore, the total number of student responses does not remain constant across each item within a course.

Differences in responses across courses for single-response items were analyzed using chi-square tests. However, as some of the possible options had a low number of student responses, Fisher's exact test was often used to determine significance. Fisher's exact test is considered more appropriate than a chi-square test when 25% of the cells of a contingency table have expected counts below 5 and a minimum expected count below 1.⁶⁶ Both chi-square and Fisher's exact tests were calculated using the stats package in the statistical software R (Version 3.6.2). Cohen's w ,⁶⁷ a measure of effect size, was calculated using the rcompanion package in R. General guidelines for Cohen's w suggest a small effect size for values around 0.1, medium around 0.3, and large around 0.5.⁶⁷ Multi-response items were also analyzed with chi-square and Fisher's exact tests. Each response option in a multi-response item was treated as a single yes-no item to create the contingency table for that option, where students who selected it were counted as 'yes' and students who did not select it were counted as 'no'. Therefore, each option within each multi-response item was analyzed as a separate contingency table to detect differences across courses. Post-hoc pairwise comparisons were only conducted when response patterns matched observed differences between course structures or features of their supporting elements (i.e., all possible pairwise comparisons were not run to hunt for significant differences). When pairwise comparisons were conducted, Fisher's exact test with a Bonferroni correction was calculated using the fmsb package in R. A significance cutoff of $p < 0.05$ for all pairwise comparisons was used unless otherwise noted.

Results and Discussion

Part 1: Structure of Face-to-Face (F2F) Learning Environment

To answer the first research question (*What are the predominant student behaviors and instructional styles for the F2F settings within these environments?*), we examined the in-class learning resources put in place by the instructors and what those meant in relation to what students were asked to "do" in the F2F sessions. The predominant student behaviors and

instructional styles identified at this stage were compared to those from the COPUS analyzer to discuss the course-to-course trends. The observations for each course, which included the COPUS codes and details of the learning activities, are compiled in the Supporting Information. The COPUS codes were compiled into a timeline that allows for the visualization of when a code was observed and for how long during each observed day of instruction, providing insight to the dynamic structure of each F2F session (Figures S1-S5).

Based on the observation of the five courses, three primary in-class resources were noted. The instructors in Courses One, Four, and Five, employed worksheets that contained problems and guiding information. The sheets were used to facilitate the majority of the F2F time and students documented their responses on the sheets. In contrast, the Course Two instructor framed the majority of the F2F time around series of clicker questions. These were used at the start of the F2F session to gauge student understanding of prior material and then throughout the session to provide real-time feedback as new content was introduced. The Course Three instructor used a combination of resources including prepared lecture slides, clicker questions, and, during one of the days, a ‘game’. Across both observation days, the majority of the F2F time was direct instruction from the prepared slides with blocks of clicker questions at the end of a module to provide formative feedback. While one of the observation days in Course Three included a group activity in the form of a game, the instructor noted that this was a deviation from the typical F2F practice. Given the flexibility in how the F2F portion of a flipped class can be supported, the range of resources utilized is not surprising. The next step was to look at how these resources were implemented and what student behaviors resulted.

A timeline of all student and instructor COPUS codes for each observation day in each course is presented in Figures S1-S5. The student codes can be grouped into behaviors where students are ‘receiving’ information, conducting ‘groupwork’, doing ‘individual work’, engaged in ‘questioning’, or doing ‘non-work’ (see Table S1 for code groupings). To compare and contrast what students were *doing* across these F2F sessions, we focused on the average amount of groupwork and questioning observed (Figure 1). With respect to groupwork, on average, more than 75% of the F2F time in Courses Four and Five included observations of students working with their peers on the instructional worksheets (WG code). In both courses, students sat in working groups of 4-5 from the start of class. In contrast, the students in Course One were observed to be engaged with their peers on the worksheets in less than 35% of the F2F time, forming groups of 2-3 each time they were directed to answer worksheet questions. Therefore, use of the same resource looked very different with regard to students’ peer-to-peer interactions over an entire F2F session. Clicker (CG) and other (OG) groupwork in Courses Two and Three was observed during roughly 50% and 40% of the F2F time respectively. Similar to the peer-to-peer interactions in Course One, clicker question discussion typically included only 2-3 students. The game in Course Three, coded as OG, involved groups of 3.

The questioning category in Figure 1 includes observations of students answering questions posed by the instructor (AnQ) or asking a question (SQ) to the instructor. These codes apply to whole-class questions not to any questions or answers during instructor facilitation of groupwork. These types of student behaviors were observed in over 80% of the F2F time in Course One and over 50% in Course Two. In both courses, the instructor used the materials (worksheet or clicker questions respectively) to introduce content and then ask whole-class

questions. In Course Two, there was a near equal balance of students responding to as well as asking whole-class questions, often times overlapping within a 2-minute time-block (displayed as AnQ & SQ in Figure 1). Student questions were infrequently observed in all other courses. In Courses Three and Four student responses to whole-class questions were observed during less than 20% of the F2F time and less than 10% in Course Five. These student behaviors were spread out in Courses Three and Five and only present during activity wrap-up sessions in Course Four.

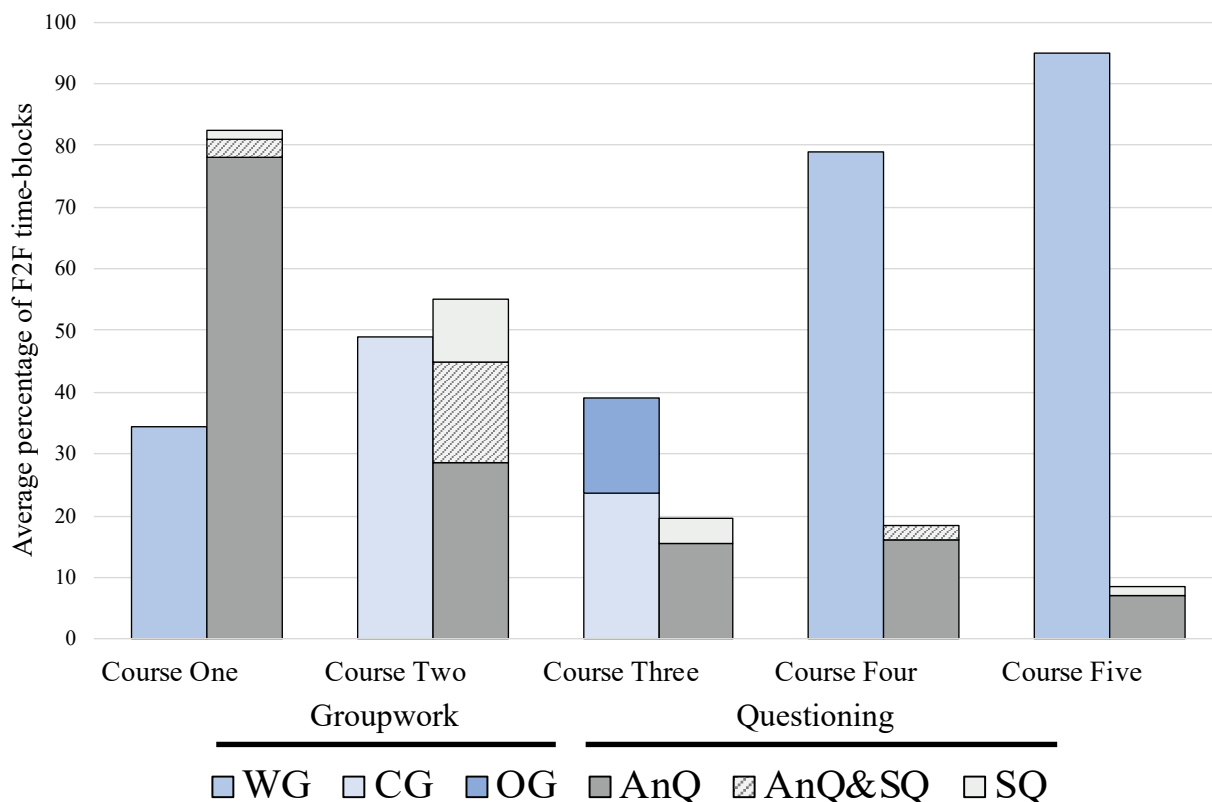


Figure 1. Average percentage of F2F time that students were observed doing ‘groupwork’ (blue bars) or ‘questioning’ (grey bars). The groupwork COPUS code category includes independent observations of ‘worksheet’ (WG), ‘clicker’ (CG), and ‘other’ (OG) activities. The questioning category includes independent observations of students ‘answering’ (AnQ) or ‘asking’ (SQ). The combined category (AnQ & SQ) notes that these two codes overlap within the time-blocks. See Table S1 in the Supporting Information for the details of each COPUS code.

The observed student behaviors in Figure 1, averaged across observation days, indicate that the predominate teaching strategy in Courses One and Two was Socratic lecturing. As noted by Freeman et. al.,⁵⁸ “Socratic lecturing involves the frequent use of questions posed to the class” with intent to “engage student attention and provide feedback to the instructor.” Course Two was balanced by the use of clicker-questions. The intent of this type of learning strategy is to develop student thinking and the application of knowledge.⁵⁸ Based on the Figure 1 groupings, the predominant teaching strategy employed in Courses Four and Five was small-groupwork on worksheets. The intent of this learning strategy is to provide students with hands-on practice with problem solving and conceptual understanding.⁵⁸ As indicated by the Course Three data in Figure 1, neither of these student behaviors dominated the F2F time. In this course, student

listening, while the instructor delivered course content from prepared slides, was also a frequently observed behavior.

COPUS Analyzer Classification of F2F Structures

The COPUS Analyzer⁶⁵ utilizes the WG, CG, OG, and SQ student codes and the lecture (Lec), posing questions (PQ), clicker questions (CG), and one-on-one discussion (1o1) instructor codes (Table S1 and Figures S1-S5) to generate clusters and associated instructional styles. When entered into the COPUS Analyzer, the observation data for each course produced three different clusters representing two different instructional styles (Table 3). The clusters for each course were consistent across the observation days.

Table 3. COPUS analyzer⁶⁵ clusters and related instructional styles.

Course	One	Two	Three	Four	Five
Cluster	6	4	4	6	5
Style	student-centered	interactive lecture	interactive lecture	student-centered	student-centered

Taken together, the COPUS timelines (Figures S1-S5) and analyzer clusters (Table 3) provide different levels of resolution for classifying each course’s F2F structure.⁶⁸ At a lower level of resolution, the identified clusters indicate the category to which each structure belongs (i.e., interactive lecture or student-centered). At a higher level, the timelines (Figures S1-S5) and aggregated student codes (Figure 1) provide additional insights to each F2F environment. In reviewing the timelines from Courses One and Four (Figures S1 and S4), in the light of the analyzer output (Table 3), it is noted that courses within the same cluster (i.e., cluster 6) can vary greatly in students’ behaviors. For example, students in Course One consistently answered questions (AnQ) posed by the instructor to the whole-class and less frequently worked in groups (WG) (Figure 1). Whereas in Course Four, students worked consistently in groups (WG) and only answered questions (AnQ) at the end of the F2F session, during the activity wrap-up (Figure S4). Therefore, despite their similar cluster groupings and defined style, students in these courses were observed to be interacting with their peers to very different levels. In contrast, the Course Five analyzer data also resulted in the student-centered instructional style but was associated with cluster 5. In considering their peer-to-peer interactions during F2F time, students in Course Five were most similar to those in Course Four. Therefore, while the COPUS analyzer and our aggregated student behaviors utilize the same data, they provide complementary information at different levels of detail for the study. It is important to note that these observations were from a two-day snapshot of each course within the overall term, thus conclusions are drawn based upon these observed days.

When considering peer-to-peer interactions during F2F time, groupwork dominated what students were *doing* in Courses Four and Five compared to the other three courses. This difference in the amount of peer-to-peer interactions is reflected upon when discussing some of the survey responses in the next section.

Part 2: Students’ Use and Perceptions of Pre-Class Materials (PCMs)

The final version of the survey was given in all five courses. The student response rate ranged from 19% to 85% (Table 4). Although Course Five had a low response rate, it was still included in the survey analyses as support for trends seen in Courses One through Four. Response rates were determined based on week-1 course enrollments (Table 1) and therefore

may not be reflective of the true rates at the time of survey administration. Students saw certain survey items based on their previous responses to items using skip logic and, as such, not all students were presented each item. Therefore, the number of students in each course that responded to each item is provided in each results table.

Table 4. Survey participation by course.

	Course One	Course Two	Course Three	Course Four	Course Five
Survey responses, n (%)	65 (36)	34 (57)	240 (85)	278 (84)	59 (19)

Do Students Watch (and Re-Watch) the Videos?

As each instructor regularly assigned videos related to the course material for each F2F session, the first survey item presented to students was, “**How many of the assigned videos have you watched?**” As can be seen in Figure 2, response distributions varied by course, from a majority responding that they watched ‘Most’ or ‘Some’ of the videos (Courses One and Two), to a majority of respondents noting that they watched ‘All’ the videos (Courses Four and Five). A chi-square test was used to determine whether there was a significant difference between the proportions of student responses by course. The result showed there was a statistically significant difference ($\chi^2(12) = 144.169, p < 0.001$); the Fisher’s exact test also showed a statistically significant difference ($p < 0.001$). The value of Cohen’s w ($w = 0.46$), suggested that this difference represents a large effect.

As course structure has been noted to influence student behavior^{52, 54} and the dominant student behavior varied across courses (Figure 1), post-hoc pairwise comparisons were conducted to detect where course-to-course differences were significant. These results showed that there was no significant difference between Courses One and Two or between Courses Three and Four. However, all other pairwise comparisons showed a significant difference, with effect sizes ranging from small ($w = 0.20$ between Courses Four and Five) to large ($w = 0.74$ between Courses One and Five). Values for all significant pairwise effect sizes are included in the description of Figure 2. The difference between courses in the percentage of students who selected that they watched ‘All’ the videos suggests that students’ viewing behaviors may trend with the structure of the in-class environment. That is, students in courses where the predominant student behavior was responding to instructor-posed whole-class questions were less likely to report watching all of the videos provided as PCMs compared to those in courses dominated by peer-to-peer interactions during groupwork. Viewing differences are also noted to align with the frequency of F2F sessions (Table 1), where Courses One and Two meet for F2F sessions three-times-per-week compared to meeting twice-weekly (Courses Three and Four) and one time only (Course Five). Viewing differences are not explained by differences in point-based incentives, as Course Five is the only one that does not provide such an incentive (Table 2).

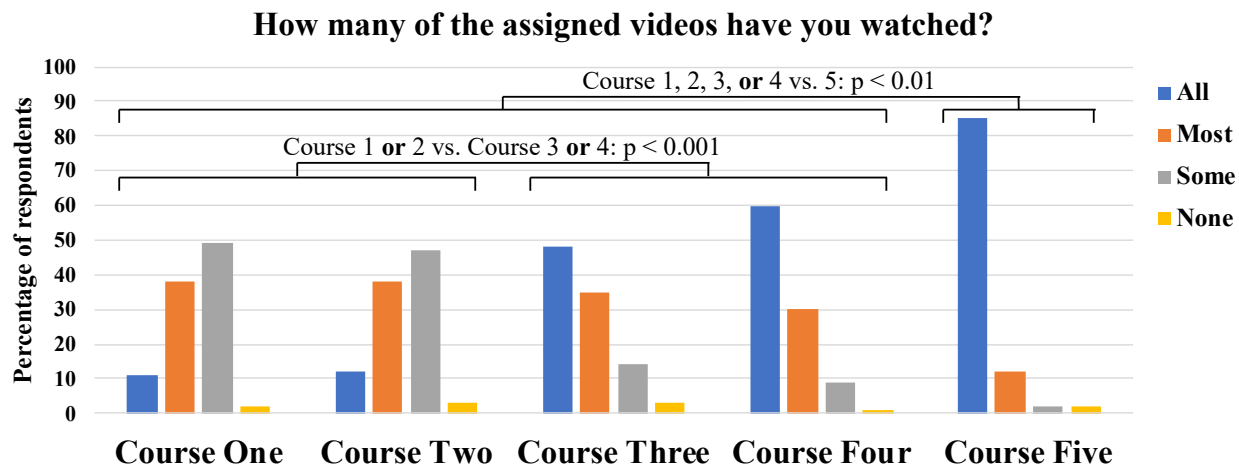


Figure 2. Percentage of respondents in each course categorized by viewing frequency. ^aSignificant pairwise comparisons between Course Five and Courses One ($w = 0.74$), Two ($w = 0.71$), Three ($w = 0.29$), or Four ($w = 0.20$). ^bSignificant pairwise comparisons between Courses One or Two and Courses Three ($w = 0.31$ and $w = 0.24$, respectively) or Four ($w = 0.39$ and $w = 0.30$, respectively).

Students who reported watching ‘Most’ or ‘Some’ of the videos were directed to respond to the item, “**Why have you not watched all of them?**” The responses from each course are provided in Table S5, found in the Supporting Information. Response options to this item were categorized into ‘General excuses’, ‘Not helpful’, and ‘I prefer other [types of resources]’. Overall, student responses to many of the options were fairly consistent across courses with a few that differed significantly. Post-hoc pairwise comparisons did not reveal any notable trends based on course structure. One potential trend of interest was based on the properties of the videos themselves. For example, Courses Four and Five had videos with longer individual run times (15+ minutes) than the other courses (Table 2). Student responses to the option ‘they [the videos] are too long’ (within the General excuses category of Table S5) differed significantly ($p < 0.001$, $w = 0.24$) with Courses Four and Five having higher percentages of students who selected this option. However, pairwise comparisons only supported a response difference for Course Four compared to Courses Two and Three, as few students in Course Five were presented with this item based on skip logic.

Because students in each course had access to the videos for the entire term after they were posted, those that watched at least ‘Some’ of them were also asked to respond to the item, “**Have you ever watched a video (or part of a video) more than once?**” Overall, a large percentage of students from each course responded that they did watch a video (or part of a video) more than once (Figure 3). These results are similar to a previous study that found students reported watching each pre-lecture video approximately three times on average.²¹ When a chi-square test was conducted, there was no statistically significant difference across courses ($\chi^2(4) = 7.8755$, $p = 0.096$).

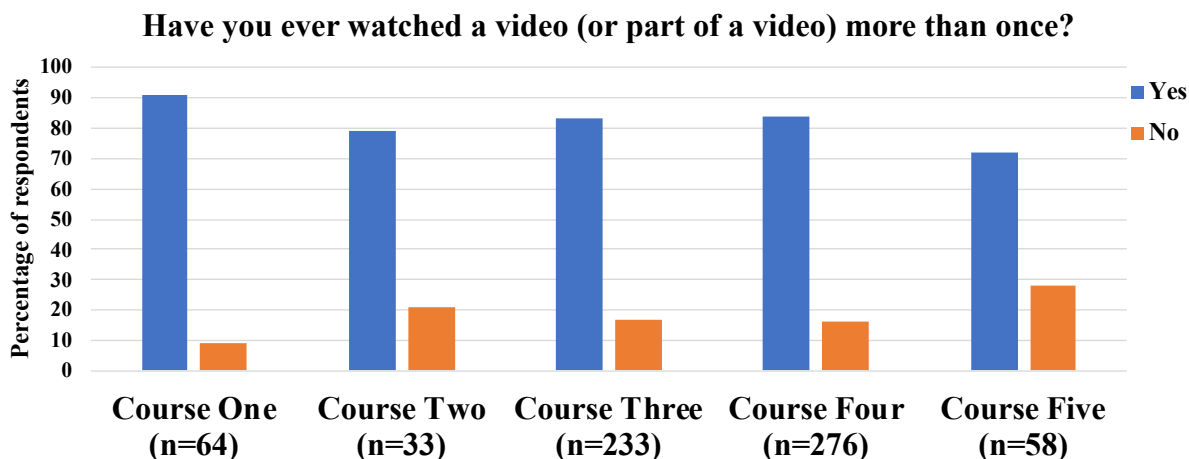


Figure 3. Percentage of respondents in each course categorized by re-watching.

Students who responded that they had never re-watched a video (Figure 3) were asked to select reasons why they had not (Table S6, Supporting Information). While the numbers of students who were presented with this follow-up item was low, the majority of respondents in most courses selected the response ‘I refer to the notes I take the first time I watch’. The exception to this were the majority respondents from Course One who selected ‘I watch other videos to get a different perspective than the ones posted’, which may be a result of the video properties itself, since Course One was the only course not supported by instructor-created screencasts (Table 2).

When Do Students Watch the Videos?

All students who responded that they watched at least ‘Some’ of the videos (Figure 2) were asked, “**When do you typically watch the videos for the first time?**” Students could only select the one option that best represented when their first viewing occurred. The percentage of student responses to this item for each course is presented in Table 5. Both chi-square ($\chi^2(16) = 171.410$, $p < 0.001$, $w = 0.51$) and Fisher’s exact tests revealed a statistically significant difference by course ($p < 0.001$). Post-hoc pairwise comparisons between the number of students who responded that they watched the videos for the first time ‘BEFORE the material is covered in class or on the homework’ showed statistically significant differences with varying effect sizes (Table 5). As a result, courses dominated by instructor-peer interaction during the F2F time (i.e., Courses One and Two), as identified by the COPUS data (Figure 1), had lower percentages of students who reported watching the videos for the first time ‘BEFORE the material was covered in class or on the homework’ when compared individually to courses with more F2F time spent in peer-to-peer interactions (Courses Four and Five). Student responses on initial viewing in Courses One and Two was split between before and after material coverage and when struggling on homework. While, to our knowledge, no earlier studies have explicitly asked students about when they viewed the videos, some have tracked behaviors through their hosting platform.^{8, 23, 31, 33, 34, 43, 69} For example, with this type of tracking information, Seery⁶⁹ reports viewings that were higher during the evening prior to an F2F activity and Ranga⁴³ reports relatively low views across the semester with spikes just prior to exams. Additionally, Bancroft et. al.²³ report a steady decline in PCM completion over a 14-week term.

Table 5. Response percentages, by course, to survey item “When do you typically watch the videos for the first time?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	64	32	233	276	58
^aWhen do you typically watch the videos for the first time? Choose the BEST option.					
Response options	Percentage of student responses to item				
BEFORE the material is covered in class or on the homework	36 ^b	28 ^b	74 ^c	90	84
AFTER the material is covered in class or on the homework	20	34	12	4	5
When I don’t understand something on a homework problem	33	28	7	2	0
When I start studying for an exam	11	9	7	3	10

^aSurvey item presented to everyone except those who selected ‘None’ for initial question, shown in Figure 2.

^bSignificant pairwise comparisons ($p < 0.05$) between Courses One or Two and Courses Three ($w = 0.33$ and $w = 0.32$, respectively), Four ($w = 0.54$ and $w = 0.49$, respectively), and Five ($w = 0.52$ and $w = 0.56$, respectively).

^cSignificant pairwise comparison ($p < 0.05$, $w = 0.22$) to Course Four.

To further understand the timing of students’ viewing, those who had responded to re-watching the videos (Figure 3) were also directed to the item, “**When do you re-watch videos?**” Response options to this item and student selections are presented in Table S7 (Supporting Information). Generally, more students responded that they would re-watch parts of a video than those that responded they would re-watch an entire video. The majority of responses across all courses for re-watching only part of a video were, ‘when I have missed something the first time’ and ‘when I need clarification at a later time (e.g., for homework or when completing a lab)’. Lower percentages reported re-watching ‘when studying for an exam’. Although some of the response options had significantly different percentages of students that selected them, no obvious trends were seen in the responses based on the F2F structure.

How Do Students Interact with the Videos?

Students who regularly watch the videos can interact with them in different ways. To assess students’ interactions with the videos, those that watched at least ‘Some’ of them (Figure 2) were asked, “**When you watch the videos, how do you watch them?**” Percentages of student responses to this item for each course are included in Table S8 (Supporting Information). Response options were categorized into ‘Pacing of viewing’ (e.g., pausing and/or rewinding) and ‘Blocking of viewing’ (e.g., all in one sitting). Overall, although there were some significant and non-significant differences between courses for the ‘Pacing of viewing’ options, no notable trends were present. The ‘Blocking of viewing’ options showed significant differences between courses. Specifically, Courses Four and Five had larger percentages of students who selected ‘I watch the assigned videos in one sitting’ instead of ‘I spread out watching the assigned videos throughout the day or week.’ Post-hoc pairwise comparisons showed a significant difference in the number of students who selected this option from Courses One, Two, and Three when compared to Courses Four and Five, ($p < 0.001$) with effect sizes ranging from small to large (details provided in Table S8). Although these differences could be influenced by the increased peer-to-peer interaction in Courses Four and Five, it could also be affected by the longer video lengths (Table 2) or the lower meeting schedule (1-2 times per week vs. 3 times per week) of those courses (Table 1).

Since the videos are meant to provide information in lieu of traditional lecture, simply watching a video without doing anything else could be considered akin to simply sitting in lecture without taking notes. Therefore, students were asked, “**When you watch the videos, what do you do while watching?**” One way to categorize students’ reported interactions with the videos is through the ICAP framework,⁷⁰ which defines certain student behaviors as indicative of Interactive, Constructive, Active, or Passive engagement. It has been found that students who interact with an activity (including content delivery) at a higher mode of engagement score higher on knowledge assessments than those who interact with the same activity at a lower mode of engagement.⁷⁰ Framed with regard to content delivery (see Table 1 in Chi et al.⁷⁰), the modes of the ICAP framework start with Passive as the lowest mode, where students are simply receiving information (e.g., listening to a lecture and not taking notes). Active is the second mode and describes students who are repeating the delivered information (e.g., copying problem solutions or taking verbatim notes). The third mode is Constructive and includes generating new information based on what is presented (e.g., solving problems, comparing and contrasting ideas, or drawing trends that were not presented). The highest mode is Interactive, which is when students participate in constructive dialoguing regarding the information. Thus, what students are doing when they watch the videos can provide some information about how they are engaging with them. The response options provided to the students that were generated from the qualitative focus group data mirror behaviors that are described in the ICAP framework.⁷⁰ Using this framework, the response options presented in Table 6 were categorized into ‘Passive’, ‘Active’, and ‘Constructive’ engagement behaviors. Because a flipped course structure has students watch the videos outside of class, there were likely no opportunities that allowed the students to participate in ‘Interactive’ engagement. One of the response options, ‘I work the problems as they are presented in the video’ could have been categorized as ‘Active’ or ‘Constructive’ engagement depending on if students were simply copying down the problems or working them out on their own. This option was categorized as ‘Constructive’ for these survey results as qualitative data from short-answer responses and focus groups indicated students would generally do the latter. This can be seen in one of the short-answer student responses collected during survey development, “*if the video contains practice problems, I pause the video and do the problem myself first.*”

Table 6. Response percentages, by course, to survey item “When you watch the videos, what do you do while watching?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	64	32	232	273	58
^aWhen you watch the videos, what do you do while watching? (Select all that apply)					
Response categories and options	Percentage of student responses to item				
<i>Passive behaviors</i>					
I just focus on the video itself (i.e., just listen or watch doing nothing else) ^{***}	44	41	28	36	5
<i>Active behaviors</i>					
I take notes on the material presented ^{***}	73	44	72	79	98
<i>Constructive behaviors</i>					
I work the problems as they are presented in the video ^{***}	73	47	63	43	67
I work on the homework problems ^{***}	27	19	21	8	5

I work on my chemistry lab assignments***	2	22 ^b	1	3	2
<i>Distracted behaviors</i>					
I do other (non-chemistry-related) activities	2	6	7	5	5

^aSurvey item presented to everyone except those who selected ‘None’ for initial question, shown in Figure 2. *** $p < 0.001$. ^bSignificant pairwise comparisons ($p < 0.05$) between Course Two and Courses One ($w = 0.35$), Three ($w = 0.35$), Four ($w = 0.27$), and Five ($w = 0.34$).

Although the percentages of students who selected the ‘Passive’ engagement response is relatively high in some courses (Table 6), it is not the majority response in any course and does not necessarily mean students were not engaging with the material at some point. As was noted earlier, a majority of students report re-watching the videos (Figure 3), usually focusing on parts that they missed or did not understand (Table S7, Supporting Information). Since students were instructed to select all the options that apply to what they do while watching, it is possible that they passively engaged the first time they watched and then actively or constructively engaged during a later time. The ability to rewind and re-watch allows students to interact with the videos with multiple modes of engagement, such that students who responded that they ‘just focus on the video itself’ could also have interacted with the video at a higher mode of engagement as well. This process was described by students in some of the focus groups that were conducted. For instance, one student stated that, *“I’ll watch it once. Not once, but I’ll watch one part through. Like if he’s doing an example, he explains a lot, I’ll just sit there and watch him explain the whole example. And then I’ll go back and pause it and just write everything down. I’m not writing and listening...that’s too much. I’ll just watch it and then write it down.”*

The difference in percentage of student responses for each response option (Table 6) when evaluated across courses were statistically significant ($p < 0.001$) with small to medium effect sizes. Only one potentially course-specific trend was noted. Course Two had a higher percentage of students who responded to the option, ‘I work on my chemistry lab assignments’, which was found to be statistically significant from the other courses with medium effect sizes when pairwise comparisons were completed. This was the only course that had the laboratory component incorporated as a part of the overall class grade, and as such, this option may only be relevant in courses where the lab is well-aligned with the course material. Lastly, course responses were low and found to not differ on the ‘Distracted behavior’ option, ‘I do other (non-chemistry-related) activities’. Overall, taking into account that students could (and often did) select more than one response option, it is difficult to say that the different percentages indicated that students in some courses were more or less engaged in the videos than in other courses.

Why Do Students Find the Videos Helpful or Not Helpful for Learning?

All students who responded that they at least watched ‘Some’ of the videos (Figure 2) were asked to respond to two survey items about what they thought made the videos **“Helpful”** or **“Not Helpful”** for their learning. The response options for the **“Helpful”** items were categorized into ‘Control of learning’ or ‘Perceived usefulness’ (Table 7). Across all courses, the majority of students responded that they found the videos helpful for their learning due to their ability to control the pace and timing/location of watching them. This matches what has been reported in other studies that collected data on what students liked best about the flipped course, specifically comments related to control of learning.^{8, 19, 24, 29, 31, 33, 34, 43, 44} The students’ perceived usefulness of the videos varied across courses, with no apparent trend present by

course, although perceived usefulness has also been found to be part of what students would comment they liked about a flipped course in other studies.^{8, 21, 31, 43} While some response options showed significant differences by course, no course-specific trends were noted.

Table 7. Response percentages, by course, to survey item “Were the videos helpful to your learning? If so, in which ways were they helpful?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	64	32	231	263	57
^aWere the videos helpful to your learning? If so, in which ways were they helpful? (Select all that apply)					
Response categories and options	Percentage of student responses to item				
<i>Control of learning</i>					
I can watch them at my own pace (e.g., rewinding, pausing, fast-forwarding)	80	78	87	81	93
I can watch them where and when I want***	72	44	75	68	88
<i>Perceived usefulness</i>					
They are easy to understand or include useful explanations and/or practice problems***	58	44	60	35	75
They help to reinforce the material***	72	44	61	33	44
They show other perspectives and ways of solving problems***	47	25	28	17	14
They often contain visual representations to understand the content*	47	41	50	49	72

^aSurvey item presented to everyone except those who selected ‘None’ for initial question, shown in Figure 2. **p < 0.01. ***p < 0.001.

When asked what was “**Not helpful**” about the videos with regard to their learning (Table S9, Supporting Information), the highest response percentages fell into the category of ‘Do not meet learning expectations’. In most courses, the majority selected that they were unable to ask questions, with the next highest (and majority in Course Three) was that the videos did not contain enough practice problems. Students’ dislike of not being able to ask questions while doing PCMs has been noted in other flipped studies as well.^{19, 33, 34} Students also provided lower percentage responses across the categories of ‘Not relevant to course’, ‘Don’t hold attention’, and ‘Poor quality/disorganized’ with no discernable patterns by course. One notable course-based response was in Course One, which used non-instructor made videos. Only 12% of students in Course One selected that ‘[the videos] have a different focus than the class material’, which was one of the lowest among the courses. This reflects well for the use of curated online videos from different sources.

Conclusions

The conclusions from this multi-course investigation are framed by our research questions.

What are the predominant student behaviors and instructional styles for the F2F settings within these environments?

Timelines of COPUS codes (Figures S1-S5), aggregated student behavior codes (Figure 1), and outputs from the COPUS Analyzer (Table 3) were used to categorize the student behaviors and instructional style of each F2F session. These sources provide complementary means⁶⁸ for categorizing the predominant student behavior and instructional style of each course for the two days in which courses were observed. Courses One, Four, and Five were categorized as “student-centered” and Courses Two and Three as “interactive lecture” by the COPUS Analyzer. However, the timelines provide higher resolution and show greater variability in styles. The predominant student behavior observed in Courses One and Two were instructor-student interactions through whole-class questioning. This practice was supplemented by brief rounds of groupwork on worksheet problems in Course One and clicker-based groupwork in Course Two. Within Course Three, student behaviors included large blocks of listening during lecture delivery of material, interspersed with responses to whole-class questioning. Additional behaviors included clicker-based groupwork on both days and a group game on day one. In Courses Four and Five, the dominant student behaviors observed were group discussions on worksheets. In Course Four, groupwork was supplemented by instructor-led wrap-up sessions where students both listened to lecture and responded to whole-class questioning. In Course Five, groupwork was supplemented by intermittent questions posed by the instructor, with students responding individually. This is the first study to our knowledge to coordinate a systematic comparison of how the F2F time is used across a variety of flipped learning classrooms at different institutions. In discussing flipped classrooms more broadly, it is important to be aware of possible heterogeneity in how this learning format is implemented, especially with regard to the degree of peer-to-peer interactions.

Within flipped classroom studies, the types of observations and comparisons conducted in this study could help address two of the critical components (structural and instructional) in a fidelity of implementation framework.⁵⁰ Structural critical components include the “expected elements related to the design and organization of the program, curriculum, or practice”.⁵⁰ Therefore, details of how an overall course is organized along with high-resolution insights to how an F2F session is structured allows educators to identify where potential key features deviate from one another when comparing outcomes. Additionally, instructional critical components include the “expected participants’ behaviors during implementation of the program, curriculum, or practice”.⁵⁰ Details in this category include aspects of both instructor and student behaviors and engagement. In studies assessing the impact of active learning or other evidence-based practices,^{3, 48} the type of activity and degree of student engagement are often cited as confounding variables that can impact learning outcomes. Therefore, flipped courses with different F2F structures likely engage students to different degrees which may lead to different outcomes. As different F2F structures may engage students to different degrees, they have the potential to lead to different course outcomes.⁷⁰ Therefore, characterizing F2F time, such as we report here using the COPUS protocol, is vital to triangulating course outcomes and providing insight if difference are or are not found between implementation of the same instructional practice. The course structures compiled in this study will be incorporated into subsequent stages of our larger multi-institution study to explore the flipped learning environment’s impacts on various aspects of student motivation and performance outcomes.

What are the students' self-reported use and perceptions of the PCMs in these environments? and What associations exist between instructional style in the F2F setting and PCM utilization?

The most salient outcome from the PCM survey was the variation students reported in their degree and timing of watching the videos. Students in Courses Four and Five more often reported watching all of the assigned videos, and doing so before the related material was addressed in the F2F session, despite the fact that these videos were generally longer than those in the other courses. The predominant student behavior in the F2F portion of these two courses consisted of groupwork on problem solving worksheets. Students in courses where the predominant behavior was responding to instructor-posed whole-class questioning, and therefore engaged in less student talk, were more likely to report watching fewer of the assigned videos and doing so after the related content was covered in an F2F session. In their study on active learning classrooms, Eddy and Hogan⁵⁴ found higher reports of study time, completion of readings before class, and higher perceived importance of preparatory work for students enrolled in courses with increased structure. They explain that their observed outcome was likely due to more student accountability built into the increased structure (e.g., more student talk during F2F time and/or increased frequency of preparatory and review assignments). While neither our study nor Eddy and Hogan's were designed to uncover why these trends existed, it is recognized that classroom environments can stimulate the development of self-regulated learning⁵⁹ and that self-regulated learners are empowered to create goals, use strategies, and implement actions to meet their goals.⁶⁰ These trends and their underlying mechanisms are worth further exploration as He et. al.⁶ found that non-compliance with recommended PCM utilization partially explained the small treatment effect of their flipped course outcomes. While other studies have looked at the impact of point-based incentives on students' PCM utilization in flipped courses,^{25, 46} these data provide a potential link between students' video viewing habits and how an F2F session is structured.

While initial viewing habits differed by course (Figure 2), more consistent responses across courses were reported with regard to re-watching habits and overall viewing behaviors. A universally high percentage of students reported re-watching the content videos in each course. The ability to watch and re-watch content videos whenever a student wants has long been touted as a benefit of flipped learning.^{19, 21} These results directly support Abeysekera and Dawson's note³⁸ that PCMs help students to self-pace their learning and therefore manage cognitive load.³⁹ When watching (either initially or upon re-watching), the majority of students reported active and constructive engagement behaviors,⁷⁰ with only a few reporting distracted behaviors. The importance of the PCMs^{32, 35, 41} and efforts to increase engagement with them³⁵ have been the focus of some recent reports on flipped learning.

Across courses, students' perception of how the PCMs were helpful was relatively consistent. The majority of students in each course ($\geq 70\%$) noted that the control of learning was a helpful aspect. This included having control over the pace of learning and well as the location and timing of learning. In addition, aspects of the students' perceived usefulness of the PCMs were consistently selected across courses. Having control over and finding usefulness in learning resources are important aspects of achievement motivation.⁷¹⁻⁷³ Therefore, student reports of control and usefulness of PCMs are important for their motivation toward the flipped learning environment, which is in turn key for their learning.⁷⁴

Relatively consistent responses were found regarding why the PCMs were not helpful. The highest response category was that the videos ‘Do not meet learning expectations’, which encompassed being unable to ask questions or interact with the instructor, and/or that the videos did not contain enough practice problems. The inability to ask questions or interact with the instructor has been a concern since the earliest chemistry-based manuscripts on flipped learning.^{19, 21} Therefore, when the idea of not being able to ask questions came up in focus groups for this study, students were asked if they *typically* asked questions during their more lecture-based courses. Most of the students did not, but felt that *the opportunity* was there if they needed to. This could indicate a perceived loss of control for students in a flipped learning environment. Because we show that the inability to ask questions over the PCMs in real-time is a student concern across multiple, distinct flipped learning styles, it is compelling to look for alternatives to these in-the-moment questions. Devoting the start of each F2F session to answering remaining questions has been reported to help in some cases,²¹ but student frustration has still been reported in others.¹⁹ As a flipped learning environment likely introduces this new aspect into the learning process for many students (i.e., self-regulated preparation using PCMs), it might be beneficial to consider their expectations for this, as well as other aspects, with regard to obtaining buy-in to the environment.⁵²

Limitations

This nonexperimental research has several limitations that should be considered when interpreting the outcomes presented. Despite the coordinated processes across multiple courses, this study only includes select introductory and general chemistry courses. Therefore, outcomes may not be generalizable to flipped chemistry courses with different student populations (e.g., by level, demographics, etc.). Additionally, this study set out simply to document information about the two main tenets of the flipped courses (i.e., pre-class content delivery and use of active learning during in-class sessions). Student performance was not collected; therefore, we cannot comment on how noted differences in F2F structure or PCM use may have impacted student outcomes. This limitation is being addressed in a subsequent phase of the larger project, where a variety of motivational and performance outcomes are also being collected within each course.

With specific regard to the classroom observations, on-site visits were scheduled with each instructor. Therefore, each knew when the consecutive observations would be conducted, potentially impacting the F2F environment structure on those two days. On-site visit timing and course variations precluded the ability to observe similar topic coverage across courses. While day-to-day consistency was seen within each course, the variability of instruction between courses could not be examined at the topic level. However, during instructor interviews none noted adjusting their practices based on content. Finally, video recordings were focused on the class as a whole, capturing a majority of the class and most instructor motion. Therefore, at the student-level, COPUS codes were applied with regard to the majority behaviors observed, not to individual or group-level behaviors. While this is typical for use of the COPUS,^{61, 62, 65} it does not capture these finer-grained variations that could impact individual student course outcomes.

Data on the pre-class materials (PCMs) only represents the perceptions and opinions of the students who self-selected to participate in the focus groups and surveys conducted. Therefore, these self-reported behaviors may not reflect those of other students, especially within

courses with low participation. Future studies are encouraged to utilize the same questions and response options to better gauge the noted response trends across more environments and student populations. Within any self-response study, students' responses could be influenced by social desirability; that is, students might respond based on what would make them "look best". However, as no data from this study was collected within the authors' institutions and none of the instructors were involved in the data collection process, the influence was potentially diminished as the research team had no connections to the students. Lastly, the PCM survey was administered midway through each course, therefore students should have been calibrated to the structure of the course and had several forms of feedback regarding their abilities and performance. This one-time survey does not account for any changes in students' use of the videos or differences in their perceptions over time. Each would be informative to capture a fuller picture of students' engagement with this aspect of a flipped learning environment.

Finally, while PCM use was observed to trend with the structure of the F2F environments, we cannot rule out other influences that may have contributed. As noted earlier, student buy-in may impact the degree to which students engage with course elements.⁵² Additionally, the consistency of student reminders has been shown to impact student use of PCMs.⁴⁶ Therefore, as these aspects were not explicitly measured within this study, they cannot be ruled out. As many flipped learning environments incentivize student use of the PCMs through on-line or in-class quizzes, further investigation of the influence of both F2F structure and incentivization may be warranted.

Implications

For practice

Although the variability between the courses included in this study does not allow for a definitive link to be made between the F2F structure and students' use of PCMs, the results provide some insight into the possible connection between the two tenets of flipped learning. When students were expected to spend a larger percentage of F2F time working in a group instead of participating in Socratic dialogue with the instructor, more of the students reported watching all of the videos and did so before the F2F session. Thus, if an instructor's expectation for students is that they watch all the videos before coming to class, it may be helpful to not only make these expectations clear to the students, but also to include more peer-to-peer groupwork in the F2F environment. If student "accountability" is a driver of increased PCM engagement, instructors may consider other ways this could be instilled, noting that the use of viewing incentive quizzes was not found to make a difference in this study. Likely unrelated to accountability, the use of regular email reminders combined with online homework has been shown to promote high levels of student engagement with PCM videos.⁴⁶

It can often be difficult for instructors to gauge the amount of time that students are actually spending doing things like engaging in peer-to-peer interactions, responding to whole-class questions, or simply listening to a presentation, which is why collecting observational data using a protocol like COPUS is essential for understanding a learning environment.⁶⁸ If COPUS results of student behaviors do not align with an instructor's expectations for the environment, then F2F time could be adjusted to account for discrepancies. In addition to documenting student behaviors, measuring students' cognitive engagement could provide another level of detail into what students are doing during F2F time. Use of the ICAP framework⁷⁰ of cognitive engagement

has found that students who “passively” engage with course material generally perform worse on knowledge assessments than students who interact with the material at a higher mode of engagement (i.e., “active”, “constructive”, or “interactive”). Even though timelines of students’ behaviors (e.g., COPUS) may indicate that students are participating in peer-to-peer interactions, these do not necessarily mean that they are cognitively engaging with them at a higher mode. Therefore, evaluating cognitive engagement can provide a deeper understanding of what students are doing in the F2F environment of a flipped course. This level of informative feedback can further influence how changes to instructional practice are implemented.

Many previous studies on flipped courses have presented results about what students thought were and were not helpful about the PCMs and the results of this study are similar to what was found previously.^{8, 19, 21, 24, 29, 31, 33, 34, 43, 44} Specifically, that students generally found that the videos were helpful based on inherent properties, such as being able to pause and rewind or being able to watch when it was best for their schedules. Students also pointed out that they found the inability to ask questions, as well as a ‘lack’ of practice problems, made the videos less helpful for their learning. As such, instructors may want to consider how best to address these aspects when implementing a flipped course. A potentially novel result found in this study was that students’ responses to what was and was not helpful about the videos did not appear to be affected by the source of the video, as Course One used non-instructor made videos curated from online sources. Although the source of the videos was not the focus of this study and further research should be done on the effect of non-instructor made videos on students’ perceptions and use of the videos, this result suggests that thoughtfully selected online videos that align with the class material are perceived as just as helpful to students as instructor-made videos.

For research

The goal of this study was to employ a coordinated set of assessment practices to evaluate the F2F environment and students’ perceptions and use of PCMs across multiple flipped chemistry courses. Subsequent phases of this project will utilize these data to triangulate course outcomes. Although previously published studies on flipped courses have been completed, they have typically focused on single courses or institutions and outcome comparisons across these studies can be limited by inconsistent assessment practices. When data is collected using coordinated assessments from multiple courses, comparisons between the courses can be directly evaluated, allowing for general trends and features to be detected and explored. We therefore encourage other researchers studying flipped classrooms, or any evidence-based instructional practice, to begin to design larger coordinated studies that may bring novel insights to our understanding of how these practices are adapted and what impacts an adaptation may have on student and course outcomes.

As variability exists in how instructors’ structure and support the two tenets of flipped learning, it is important to provide information about PCM use and the F2F environment structure when presenting results. Without implementation details, the validity of findings from a flipped learning environment may be compromised.⁵⁰ The F2F observations made in this study included multiple levels of information, such as the COPUS analyzer⁶⁵ clusters, percentages of class time that students spent on different activities, and timelines of student and instructor COPUS codes. The COPUS analyzer clusters provided information about the general ‘type’ of F2F environment that was implemented in each course and provided a lower-resolution picture

of the F2F style. Additional data from the percentages of class time students spent on different activities provided more detail into what students were doing in the F2F environment, as this was observed to differ even in courses that were part of the same cluster. Taken together, these data provided a general picture of the F2F environment at a similar degree of resolution as observations made by Cannelas et al.,¹¹ who reported the percentage of class time the students spent being “active” (i.e., everything except watching, listening, and taking notes). The COPUS timelines presented higher-resolution data, specifically, what was happening in the classroom every 2 minutes. Timelines allow for more details to be presented regarding when different behaviors occur in the classroom. For instance, a study by Donnelly and Hernández²⁶ presented the percentage of students engaged during the F2F portion of a flipped classroom for each 2-minute interval and found that student engagement fluctuated throughout class time and type of active learning (i.e., whole-class discussion vs. group activity). Therefore, although details of the course’s structure should always be included in a study, deciding to include and/or emphasize either the low- or high-resolution data about the F2F environment should be dependent on the specific research question being asked.

Finally, although this study explored the relation between the F2F environment and students’ perceptions and use of PCMs in flipped chemistry classrooms, data about every facet of the environment was not gathered. Therefore, continued research into these two tenets of flipped courses is needed to better understand how students’ behaviors, engagement, and learning is affected by these environments. Qualitative studies with student focus groups and/or interviews could be used to ask students about why they do or do not engage in different behaviors with the PCMs and in the F2F environment. Additionally, qualitative or mixed-methods studies could be coupled with tracking data to gather more details about if and when students are using PCMs based on the expectations of the F2F environment or levels or assessments and if these behaviors change throughout the course of the term. Further investigation into “if”, “when”, and “why” students do or do not engage in these two tenets could provide valuable information about why outcome differences are seen between different flipped course environments.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.XXXXXXX. [ACS will fill this in.] COPUS code descriptions, F2F observation summaries, COPUS timelines, survey development details, supplementary survey item data tables (DOCX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: jack.barbera@pdx.edu

ORCID

Jack Barbera: [0000-0003-3887-3301](https://orcid.org/0000-0003-3887-3301)

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. (DUE 1611220 and 1611519). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

This project would not have been possible without the participation of each instructor and the students in their flipped courses, we are therefore very thankful for their assistance in this study. Classroom observations and instructor interviews were collected by Regis Komperda, she is thanked for her time and willingness to travel to each institution. Cory Hensen is thanked for taking the lead on the initial focus group coding and for training and mentoring the secondary coders. Carrie Zografos, Andrew Isom, and Gosia Cox are thanked for their assistance in COPUS coding. Nicole James is thanked for her assistance in structuring and editing the manuscript itself.

References

1. President's Council of Advisors on Sciences and Technology (PCAST) *Engaged to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*; Washington, DC, 2012.
2. National Research Council *Discipline-based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; Washington, DC, 2012.
3. 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 USA* **2014**, *111* (23), 8410-5.
4. Shadle, S. E.; Marker, A.; Earl, B., Faculty drivers and barriers: laying the groundwork for undergraduate STEM education reform in academic departments. *Int J STEM Educ* **2017**, *4* (1), 8.
5. Bergmann, J.; Sams, A., *Flip Your Classroom: Reach every student in every class every day*. International Society for Technology in Education: Eugene, OR, 2012.
6. He, W.; Holton, A.; Farkas, G.; Warschauer, M., The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learn. Instr.* **2016**, *45*, 61-71.
7. Liu, Y.; Raker, J. R.; Lewis, J. E., Evaluating student motivation in organic chemistry courses: moving from a lecture-based to a flipped approach with peer-led team learning. *Chem. Educ. Res. Pract.* **2018**, *19* (1), 251-264.
8. Mooring, S. R.; Mitchell, C. E.; Burrows, N. L., Evaluation of a Flipped, Large-Enrollment Organic Chemistry Course on Student Attitude and Achievement. *J. Chem. Educ.* **2016**, *93* (12), 1972-1983.
9. Rein, K. S.; Brookes, D. T., Student Response to a Partial Inversion of an Organic Chemistry Course for Non-Chemistry Majors. *J. Chem. Educ.* **2015**, *92* (5), 797-802.
10. Robert, J.; Lewis, S. E.; Oueini, R.; Mapugay, A., Coordinated Implementation and Evaluation of Flipped Classes and Peer-Led Team Learning in General Chemistry. *J. Chem. Educ.* **2016**, *93* (12), 1993-1998.
11. Canelas, D. A.; Hill, J. L.; Novicki, A., Cooperative learning in organic chemistry increases student assessment of learning gains in key transferable skills. *Chem. Educ. Res. Pract.* **2017**, *18* (3), 441-456.

12. Hibbard, L.; Sung, S.; Wells, B., Examining the Effectiveness of a Semi-Self-Paced Flipped Learning Format in a College General Chemistry Sequence. *J. Chem. Educ.* **2015**, *93* (1), 24-30.
13. Christiansen, M. A.; Nadelson, L.; Etchberger, L.; Cuch, M.; Kingsford, T. A.; Woodward, L. O., Flipped Learning in Synchronously-Delivered, Geographically-Dispersed General Chemistry Classrooms. *J. Chem. Educ.* **2017**, *94* (5), 662-667.
14. Flynn, A. B., Structure and evaluation of flipped chemistry courses: organic & spectroscopy, large and small, first to third year, English and French. *Chem. Educ. Res. Pract.* **2015**, *16* (2), 198-211.
15. Hill, D. J.; Williams, O. F.; Mizzy, D. P.; Triumph, T. F.; Brennan, C. R.; Mason, D. C.; Lawrence, D. S., Introduction to Laboratory Safety for Graduate Students: An Active-Learning Endeavor. *J. Chem. Educ.* **2019**, *96* (4), 652-659.
16. Bernard, P.; Broś, P.; Migdał-Mikuli, A., Influence of blended learning on outcomes of students attending a general chemistry course: summary of a five-year-long study. *Chem. Educ. Res. Pract.* **2017**, *18* (4), 682-690.
17. Blackburn, R. A. R., Write My Next Lecture: Prelecture Problem Classes and In-Lecture Discussion To Assist Case-Study Teaching of Synthesis. *J. Chem. Educ.* **2017**, *95* (1), 104-107.
18. Bokosmaty, R.; Bridgeman, A.; Muir, M., Using a Partially Flipped Learning Model To Teach First Year Undergraduate Chemistry. *J. Chem. Educ.* **2019**, *96* (4), 629-639.
19. Ealy, J. B., Development and Implementation of a First-Semester Hybrid Organic Chemistry Course: Yielding Advantages for Educators and Students. *J. Chem. Educ.* **2013**, *90* (3), 303-307.
20. 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.* **2015**, *93* (1), 13-23.
21. Smith, J. D., Student attitudes toward flipping the general chemistry classroom. *Chem. Educ. Res. Pract.* **2013**, *14* (4), 607-614.
22. Amaral, K. E.; Shank, J. D.; Shibley, I. A.; Shibley, L. R., Web-Enhanced General Chemistry Increases Student Completion Rates, Success, and Satisfaction. *J. Chem. Educ.* **2013**, *90* (3), 296-302.
23. Bancroft, S. F.; Fowler, S. R.; Jalaeian, M.; Patterson, K., Leveling the Field: Flipped Instruction as a Tool for Promoting Equity in General Chemistry. *J. Chem. Educ.* **2019**.
24. Christiansen, M. A., Inverted Teaching: Applying a New Pedagogy to a University Organic Chemistry Class. *J. Chem. Educ.* **2014**, *91* (11), 1845-1850.
25. Christiansen, M. A.; Lambert, A. M.; Nadelson, L. S.; Dupree, K. M.; Kingsford, T. A., In-Class Versus At-Home Quizzes: Which is Better? A Flipped Learning Study in a Two-Site Synchronously Broadcast Organic Chemistry Course. *J. Chem. Educ.* **2016**, *94* (2), 157-163.
26. Donnelly, J.; Hernández, F. E., Fusing a reversed and informal learning scheme and space: student perceptions of active learning in physical chemistry. *Chem. Educ. Res. Pract.* **2018**, *19* (2), 520-532.
27. Fautch, J. M., The flipped classroom for teaching organic chemistry in small classes: is it effective? *Chem. Educ. Res. Pract.* **2015**, *16* (1), 179-186.
28. Fitzgerald, N.; Li, L., Using Presentation Software To Flip an Undergraduate Analytical Chemistry Course. *J. Chem. Educ.* **2015**, *92* (9), 1559-1563.
29. Gregorius, R. M., Performance of underprepared students in traditional versus animation-based flipped-classroom settings. *Chem. Educ. Res. Pract.* **2017**, *18* (4), 841-848.

30. Lenczewski, M. S., Scaffolded Semi-Flipped General Chemistry Designed To Support Rural Students' Learning. *J. Chem. Educ.* **2016**, *93* (12), 1999-2003.
31. Parsons, A. F., Flipping Introductory Retrosynthetic Analysis: An Exemplar Course To Get the Ball Rolling. *J. Chem. Educ.* **2019**, *96* (4), 819-822.
32. 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.
33. Shattuck, J. C., A Parallel Controlled Study of the Effectiveness of a Partially Flipped Organic Chemistry Course on Student Performance, Perceptions, and Course Completion. *J. Chem. Educ.* **2016**, *93* (12), 1984-1992.
34. Weaver, G. C.; Sturtevant, H. G., Design, Implementation, and Evaluation of a Flipped Format General Chemistry Course. *J. Chem. Educ.* **2015**, *92* (9), 1437-1448.
35. 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.
36. Kavanagh, L.; Reidsema, C.; McCredden, J.; Smith, N., In *The Flipped Classroom: Practice and Practices in Higher Education*, Reidsema, C.; Kavanagh, L.; Hadgraft, R.; Smith, N., Eds. Springer Nature: Singapore, 2017; pp 15-35.
37. Pienta, N. J., Introductory Chemistry Using the "Flipped" Environment: An Update. *J. Chem. Educ.* **2019**, *96* (6), 1053-1054.
38. Abeysekera, L.; Dawson, P., Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *High. Educ. Res. Dev.* **2014**, *34* (1), 1-14.
39. Clark, R. C.; Nguyen, F.; Sweller, J., *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. Pfeiffer: San Francisco, CA, 2005.
40. Sweller, J., Cognitive load theory, learning difficulty, and instructional design. *Learn. Instr.* **1994**, *4* (4), 295-312.
41. Casselman, M. D.; Atit, K.; Henbest, G.; Guregyan, C.; Mortezaei, K.; Eichler, J. F., Dissecting the Flipped Classroom: Using a Randomized Controlled Trial Experiment to Determine When Student Learning Occurs. *J. Chem. Educ.* **2019**.
42. Moreno, R.; Mayer, R. E., Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology* **1999**, *91* (2), 358-368.
43. Ranga, J. S., Customized Videos on a YouTube Channel: A Beyond the Classroom Teaching and Learning Platform for General Chemistry Courses. *J. Chem. Educ.* **2017**, *94* (7), 867-872.
44. Reid, S. A., A flipped classroom redesign in general chemistry. *Chem. Educ. Res. Pract.* **2016**, *17* (4), 914-922.
45. Webber, D. M.; Flynn, A. B., How Are Students Solving Familiar and Unfamiliar Organic Chemistry Mechanism Questions in a New Curriculum? *J. Chem. Educ.* **2018**, *95* (9), 1451-1467.
46. Woodward, R. L.; Reid, C. S., You've Got Mail (and Homework): Simple Strategies for Promoting Student Engagement with Prelecture Videos. *J. Chem. Educ.* **2019**, *96* (9), 2055-2058.
47. Pilcher, S. C., Hybrid Course Design: A Different Type of Polymer Blend. *J. Chem. Educ.* **2017**, *94* (11), 1696-1701.
48. Rahman, T.; Lewis, S. E., Evaluating the evidence base for evidence-based instructional practices in chemistry through meta-analysis. *J. Res. Sci. Teach.* **2019**.

49. Seery, M. K., Flipped learning in higher education chemistry: emerging trends and potential directions. *Chem. Educ. Res. Pract.* **2015**, *16* (4), 758-768.
50. Stains, M.; Vickrey, T., Fidelity of Implementation: An Overlooked Yet Critical Construct to Establish Effectiveness of Evidence-Based Instructional Practices. *CBE Life Sci Educ* **2017**, *16* (1).
51. Wieman, C. E., Large-scale comparison of science teaching methods sends clear message. *P. Natl. Acad. Sci. USA* **2014**, *111* (23), 8319-20.
52. Cavanagh, A. J.; Aragon, O. R.; Chen, X.; Couch, A.; Durham, F.; Bobrownicki, A.; Hanauer, D. I.; Graham, M. J., Student Buy-In to Active Learning in a College Science Course. *CBE Life Sci Educ* **2016**, *15* (4).
53. Brazeal, K. R.; Couch, B. A., Student Buy-In Toward Formative Assessments: The Influence of Student Factors and Importance for Course Success. *J Microbiol Biol Educ* **2017**, *18* (1).
54. Eddy, S. L.; Hogan, K. A., Getting under the hood: how and for whom does increasing course structure work? *CBE Life Sci Educ* **2014**, *13* (3), 453-68.
55. Freeman, S.; O'Connor, E.; Parks, J. W.; Cunningham, M.; Hurley, D.; Haak, D.; Dirks, C.; Wenderoth, M. P., Prescribed active learning increases performance in introductory biology. *CBE Life Sci Educ* **2007**, *6* (2), 132-9.
56. Brown, T. L.; Brazeal, K. R.; Couch, B. A., First-Year and Non-First-Year Student Expectations Regarding In-Class and Out-of-Class Learning Activities in Introductory Biology. *J Microbiol Biol Educ* **2017**, *18* (1).
57. He, W.; Holton, A.; Gu, H.; Warschauer, M.; Farkas, G., Differentiated Impact of Flipped Instruction: When Would Flipped Instruction Work or Falter? *International Journal of Teaching and Learning in Higher Education* **2019**, *31*, 32-49.
58. Freeman, S.; Haak, D.; Wenderoth, M. P., Increased course structure improves performance in introductory biology. *CBE Life Sci Educ* **2011**, *10* (2), 175-86.
59. Zimmerman, B. J., Self-efficacy and educational development. In *Self-efficacy in changing societies*, Bandura, A., Ed. Cambridge University Press: New York, 1995.
60. Ridley, D. S.; Schutz, P. A.; Glanz, R. S.; Weinstein, C. E., Self-regulated learning: The interactive influence of metacognitive awareness and goal-setting. *The journal of experimental education* **1992**, *60* (4), 293-306.
61. Smith, M. K.; Jones, F. H.; Gilbert, S. L.; Wieman, C. E., The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci Educ* **2013**, *12* (4), 618-27.
62. Lund, T. J.; Pilarz, M.; Velasco, J. B.; Chakraverty, D.; Rosploch, K.; Undersander, M.; Stains, M., The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE Life Sci Educ* **2015**, *14* (2).
63. Stains, M.; Harshman, J.; Barker, M. K.; Chasteen, S. V.; Cole, R.; DeChenne-Peters, S. E.; Eagan, M. K., Jr.; Esson, J. M.; Knight, J. K.; Laski, F. A.; Levis-Fitzgerald, M.; Lee, C. J.; Lo, S. M.; McDonnell, L. M.; McKay, T. A.; Michelotti, N.; Musgrove, A.; Palmer, M. S.; Plank, K. M.; Rodela, T. M.; Sanders, E. R.; Schimpf, N. G.; Schulte, P. M.; Smith, M. K.; Stetzer, M.; Van Valkenburgh, B.; Vinson, E.; Weir, L. K.; Wendel, P. J.; Wheeler, L. B.; Young, A. M., Anatomy of STEM teaching in North American universities. *Science* **2018**, *359* (6383), 1468-1470.

64. Cohen, J., A coefficient of agreement for nominal scales. *Educational and psychological measurement* **1960**, *20* (1), 37-46.
65. Stains, M.; Harshman, J. COPUS Analyzer. <http://www.copusprofiles.org> (accessed January 21, 2020).
66. Mayers, A., *Introduction to statistics and SPSS in psychology*. Pearson Higher Ed: 2013.
67. Cohen, J., A power primer. *Psychological Bulletin* **1992**, *112* (1), 155-159.
68. Reisner, B. A.; Pate, C. L.; Kinkaid, M. M.; Paunovic, D. M.; Pratt, J. M.; Stewart, J. L.; Raker, J. R.; Bentley, A. K.; Lin, S.; Smith, S. R., I've Been Given COPUS (Classroom Observation Protocol for Undergraduate STEM) Data on My Chemistry Class... Now What? *J. Chem. Educ.* **2020**, *97* (4), 1181-1189.
69. Seery, M. K., ConfChem Conference on Flipped Classroom: Student Engagement with Flipped Chemistry Lectures. *J. Chem. Educ.* **2015**, *92* (9), 1566-1567.
70. Chi, M. T. H.; Wylie, R., The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educ. Psychol.* **2014**, *49* (4), 219-243.
71. Deci, E. L.; Ryan, R. M., The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry* **2000**, *11* (4), 227-268.
72. Niemiec, C. P.; Ryan, R. M., Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and research in Education* **2009**, *7* (2), 133-144.
73. Ryan, R. M.; Deci, E. L., Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* **2000**, *55* (1), 68.
74. Anderman, E. M.; Dawson, H., Learning with motivation. In *Handbook of research on learning and instruction*, Mayer, R. E.; Alexander, P. A., Eds. Routledge: New York, N.Y., 2011; pp 219-241.

Supporting Information

A Multi-course Comparative Study of the Core Aspects for Flipped Learning: Investigating In-Class Structure and Student Use of Video Resources

^aNicole Naibert, ^aEmryse Geye, ^bMichael M. Phillips, and ^{a*}Jack Barbera

^aPortland State University – Department of Chemistry, Portland Oregon

^bUniversity of Northern Colorado – School of Psychological Sciences, Greeley Colorado

*jack.barbera@pdx.edu

Table of Contents

COPUS CODE DESCRIPTIONS FOR SELECTED CODES.....	2
STUDENT CODES.....	2
INSTRUCTOR CODES.....	2
F2F OBSERVATION SUMMARIES AND COPUS TIMELINES	4
COURSE ONE.....	4
COURSE TWO	5
COURSE THREE	6
COURSE FOUR.....	8
COURSE FIVE	9
PRE-CLASS MATERIAL (PCM) SURVEY DEVELOPMENT.....	11
FOCUS GROUP PARTICIPANTS.....	11
FOCUS GROUP DATA COLLECTION AND ANALYSIS PROTOCOLS	11
FOCUS GROUP RESULTS.....	12
PILOT SURVEY POPULATION AND ADMINISTRATION	13
PILOT SURVEY DATA COLLECTION AND ANALYSIS PROTOCOLS.....	13
PILOT SURVEY RESULTS AND SURVEY REFINEMENT.....	14
SUPPLEMENTARY TABLES.....	17

COPUS Code Descriptions for Selected Codes

Student Codes

The listening (**L**) code was used across all courses to indicate any time students were expected to be listening to the instructor/presenter, with or without an expectation of note-taking, such as during lecture or follow-up activities, but also while the instructor presented administrative information.

The individual work (**Ind**) code was used for any type of independent work that was not an official test or quiz. In Course Two, this code was used when students conducted "speed drills" over prior material at the beginning of class, and in Course Three, this code referred to a pre-test and post-test worksheet related to the in-class game.

The clicker groupwork (**CG**) code was only used when a clicker question was posed to students and the students discussed the question among themselves in groups of two or more, whereas the worksheet group (**WG**) code was only used if the students had a given worksheet/workbook on which to work, whether that worksheet/workbook was physical as in Courses One and Four, or digital as in Course Five. In Course Three, the other groupwork (**OG**) code was used on day one to denote the game students played as groups, and on day two was used to denote groupwork where students were asked to answer a question that was neither part of a worksheet nor a clicker question.

The answer question (**AnQ**) student code was used across all courses when students answered non-rhetorical questions posed by the instructor with the rest of the class listening, whereas the student question (**SQ**) code was used when students posed questions to the instructor, whether subject-matter related or administrative, with the rest of the class listening. In Course Two, in addition to the usual sense, the SQ code was used while the instructor read and answered student questions submitted to the clicker input system out loud to the class.

The test or quiz (**TQ**) code was used any time students took a test or quiz that was handed in during class to be graded. As participating courses were observed only during non-exam weeks, this code only refers to an in-class quiz given in Course One. The waiting (**W**) student code was used when instructors had technical problems with their presentations and students were waiting for class to resume. The other (**O**) code was used to denote students coming into class late or leaving early, pointing out inconsistencies in the text that was presented on the board/screen, or giving general comments.

Instructor Codes

The lecture (**Lec**) code was used across all courses when the instructor presented subject-matter related information to the class that was not related to an example problem, whereas the follow-up (**FIUP**) code was used across all courses when the instructor explained and/or followed-up on a student activity, such as clicker questions, or group or individual work. The real-time writing (**RtW**) code was used any time the instructor wrote on a surface for the whole class to see, such as a whiteboard, smartboard, or document projector.

The pose question (**PQ**) instructor code was used across all courses when an instructor posed a non-rhetorical question to the whole class and gave space for individual students to

answer with the whole class listening. In Course Two, this code was also used to denote when the instructor administered "speed drills," as they were non-rhetorical and non-clicker questions posed to the entire class that the students were expected to answer. The clicker question (**CQ**) code was used any time an instructor conducted whole-class polling, with or without clicker software. For example, instructors could ask students to raise their hand in a vote, such as in Course One, or to raise colored response cards, such as in Course Five. Clicker questions could be to assess course knowledge, such as in Courses Two, Three, and Five, or to poll students regarding administrative matters. For example, Course One used a hand-raising poll to vote on when homework assignments would be due and Course Two used clicker polling for students to ask questions of the instructor or to self-report how many "speed drill" questions they answered correctly. The answer question (**AnQ**) instructor code was used across all courses when instructors answered questions posed by the students, whether subject-matter related or administrative, with the rest of the class listening. In Course Two, in addition to the usual sense, AnQ was used while the instructor read and answered student questions submitted to the clicker input system out loud to the class.

The moving-and-guiding (**MG**) code was used across all courses whenever the instructor was moving around the room and guiding work, either group or individual. In courses with learning assistants (LAs), such as Courses One, Three, and Four, this code was used if the instructor or even one of the LAs were moving through and guiding groups in a given time block, rather than indicating that all LAs were currently circulating. Relatedly, the one-on-one (**1o1**) code was used to indicate a directed conversation between the instructor with a single student or group. In courses with LAs, this code was used if the instructor or even one of the LAs participated in a one-on-one in a given time block, rather than indicating that all LAs were occupied in one-on-ones.

The administration (**Adm**) code was used when the instructor gave class-wide announcements related to facilitating the class meeting (e.g., time left on an activity) or course reminders (e.g., test dates, school closures, etc.) In addition to the usual sense, in Course Three, Adm was used when the instructor gave instructions regarding the in-class game. In Course Five, the demo (**D**) code was used to indicate the instructor performing a lab experiment in class. In Course Three, the waiting (**W**) instructor code was used when there was an opportunity for the instructor to be interacting with or observing/listening to student or group activities and the instructor was not doing so. The other (**O**) instructor code was used to denote instructors conferencing with LAs or researchers, or leaving the room entirely. In Course Three, O was used during day one to indicate the LAs collecting and/or distributing materials related to the in-class game, usually at the same time the instructor was still giving instructions (Adm) at the front of the classroom.

F2F Observation Summaries and COPUS Timelines

The COPUS protocol contains 13 student and 12 instructor categories to utilize during each two-minute interval within a F2F environment.⁵¹ These categories are meant to capture the range of common behaviors that typically occur in courses. Of the 25 codes, 22 were observed to occur across the recorded F2F sessions (Table S1). The categories not observed included those associated with student presentations, predictions and whole-class discussion.

Table S1. COPUS codes from observations, colors matched to course timelines shown in Figures S1-S5.

Student	Code	Description
Receiving	L	Listening to instructor
Groupwork	WG	Working in groups on worksheet activity
	CG	Discussing clicker question
	OG	Other assigned group activity
Individual Work	Ind	Individual thinking/problem solving
Q&A	AnQ	Answering question posed by instructor
	SQ	Asking a question
Non-work	TQ	Test or quiz
	W	Waiting
	O	Other
Instructor	Code	Description
Presenting	Lec	Lecturing or presenting information
	RtW	Real-time writing
	D/V	Showing or conducting a demo, experiment, simulation, etc.
Engaging	FIUp	Follow-up/feedback on clicker question or activity
	MG	Moving through class guiding ongoing student work
	1o1	One-on-one extended discussion with individual students
Q&A	PQ	Posing non-clicker question to students (non-rhetorical)
	CQ	Asking clicker question (entire duration)
	AnQ	Listening to and answering student questions to entire class
Non-work	Adm	Administration (assign homework, return materials)
	W	Waiting
	O	Other

For each two-minute time-block, the timelines indicate which student and/or instructor codes were observed (Figures S1-S5). For further interpretation, we also determined the percentage of time-blocks in which a specific code was observed. These values are presented at the end of each code's row in a given timeline.

Course One

Class sessions revolved around the discussion and completion of pages from an instructor-authored workbook. The instructor displayed pages of the workbook on a document camera and worked through each problem using a variety of methods including lecturing (Lec, 17% of day one and 40% of day two 2-minute time blocks shown in Figure 1), whole-class questioning (PQ, 83% and 80%) with follow-ups (FIUp, 63% and 49%), and real-time writing (RtW, 69% and 74%). The temporal associations of these often-overlapping methods is shown in Figure 1. Students were observed mainly listening (L, 94% each day) or responding to instructor-posed whole-class questions (AnQ, 83% and 80%). Groupwork (WG, 29% and 40%) was observed to be spread throughout the class time, typically lasting between 2-8 minutes. The instructor and/or learning assistants were observed moving through the class guiding each group's work (MG, 26% and 37%). When these timelines were analyzed with the COPUS

Analyzer⁵⁵, the COPUS Profile matched cluster 6, representing a student-centered instructional style.

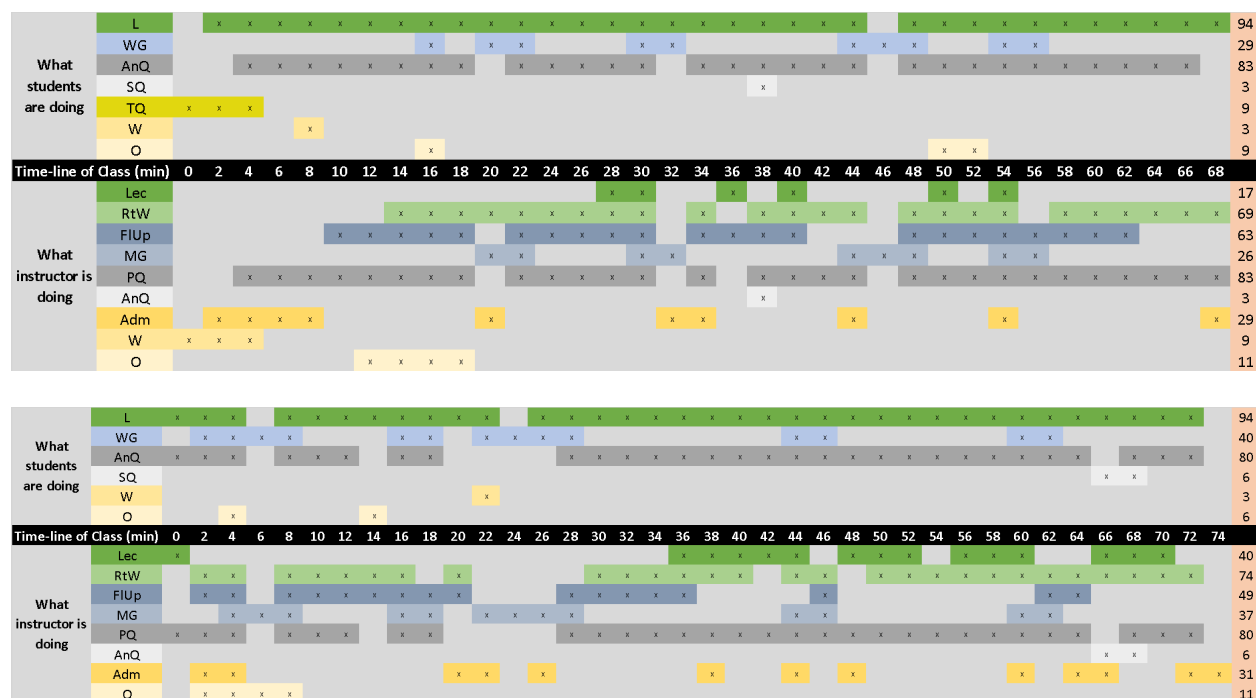


Figure S1. COPUS code timelines for Course One. Observation day one on top and day two on bottom. Code abbreviations and colors are provided in Table S1. Values at the end of each row are the percentage of 2-minute time blocks the code was observed.

The workbook used in Course One contained POGIL-style worksheets, by topic, that included models (often figures from the textbook) and text-like explanations of the content. Worksheet topics were aligned with the assigned pre-class videos. The worksheet from the day one observation contained models on the topic of periodic trends, models from day two covered ionic compound naming. Each model was followed by a series of questions. Initial questions asked students to extract and explain information from the model (e.g., key questions in the POGIL framework)⁵⁸ to acquire new knowledge. These were followed by exercises requiring students to use the information presented in the model.

Course Two

Class sessions revolved around individual- and group-based questions posed by the instructor. Each class began with a set of questions about prior material. Students were observed to work independently (Ind, Figure S2) during these “speed drills” and then entering their responses into the clicker application using their phone or tablet (12% and 13% of each day’s two 2-minute time blocks). The instructor followed up on these questions with brief clarification as needed. The remainder of the class time moved back and forth between students responding to group-based clicker questions (CG, 48% and 50%), students posing questions (SQ, 32% and 21%), and the instructor following-up on questions (FIUp, 56% and 50%). During the group clicker questions, the instructor was observed to be guiding student/groupwork (MG, 24% and 33%) and answering questions (1o1, 16% and 25%); no learning assistants were present. The majority of the class time was back-and-forth questioning by the instructor (PQ, 56% and 50%)

and answering by the students (AnQ, 44% and 46%). The COPUS Profile for these timelines matched cluster 4, indicating an interactive lecture instructional style.

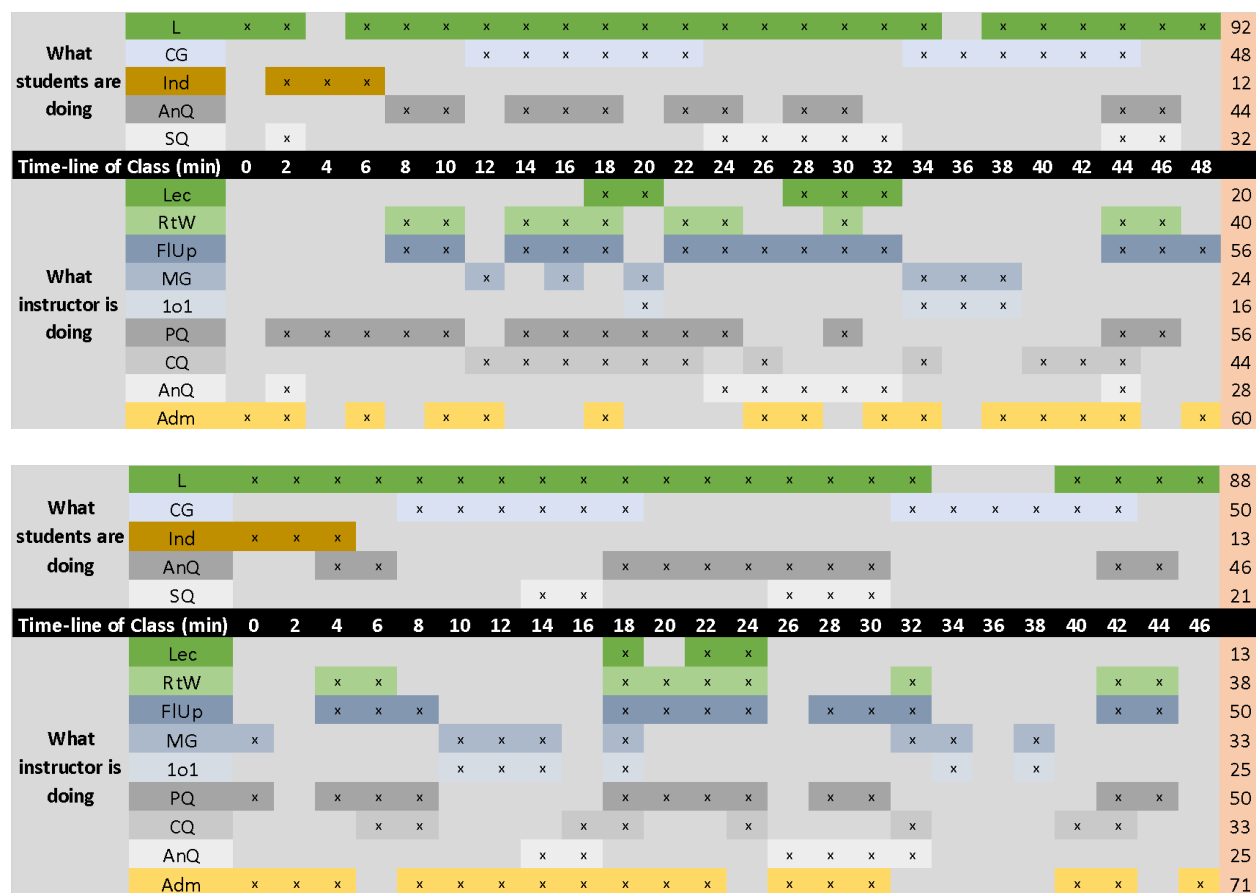


Figure S2. COPUS code timelines for Course Two. Observation day one on top and day two on bottom. Code abbreviations and colors are provided in Table S1. Values at the end of each row are the percentage of 2-minute time blocks the code was observed.

Class sessions in Course Two entailed 10-15 clicker questions presented from a tablet, allowing the instructor to annotate responses as needed. The content of each class session was aligned with the assigned pre-class videos. Day one questions covered electron transfer reactions. Initial items included a series of reaction classification questions followed by a longer series regarding the determination of the oxidation number for a selected atom in a compound. Subsequent questions asked students to apply the concept to determining the oxidized and reduced species for a specified redox reaction. Initial items on day two covered oxidation number assignments followed by a series of reaction completion questions where students were given reactants and a table of cell potentials and asked to predict the products. A final set of questions asked students to identify the number of electrons transferred or the oxidizing agent for several redox reactions.

Course Three

Class sessions revolved around cycles of topic introduction, examples, and a topic review (Figure S3). Each cycle was observed to begin with the instructor introducing a topic through lecture (Lec, 38% and 74% of each day's two 2-minute time blocks) and annotating the prepared

slides (RtW, 3% and 51%). Following this introduction, students were presented with an example to work. Students were then observed discussing with their neighbors (CG, 21% and 26%), and submitting their individual responses into the clicker application using their phone or computer. During this time, the instructor and learning assistants were observed to be guiding student/groupwork (MG, 41% and 18%). Once responses were submitted, the instructor was observed reviewing the example and providing additional context as needed (FIUp, 13% and 31%). In addition to this typical in-class practice, day one included a non-typical group activity presented in the form of a game. Midway through the class session, students individually (Ind, 15%) completed and turned in a pre-test worksheet. They were then given instructions (coded as Adm) for the game, which was based on the topic of intermolecular forces. Students worked together to play the game (OG, 28%), racing to respond to questions from an arbiter. They then completed an individual post-test worksheet. This non-typical activity accounted for the variation in code percentages across observation days, especially in the case of the administration code (Adm, 49% and 8%), used to note the instructor’s explanations of the game’s rules and directions regarding the worksheets. The COPUS Profile for these timelines matched an interactive lecture instructional style represented by cluster 4.

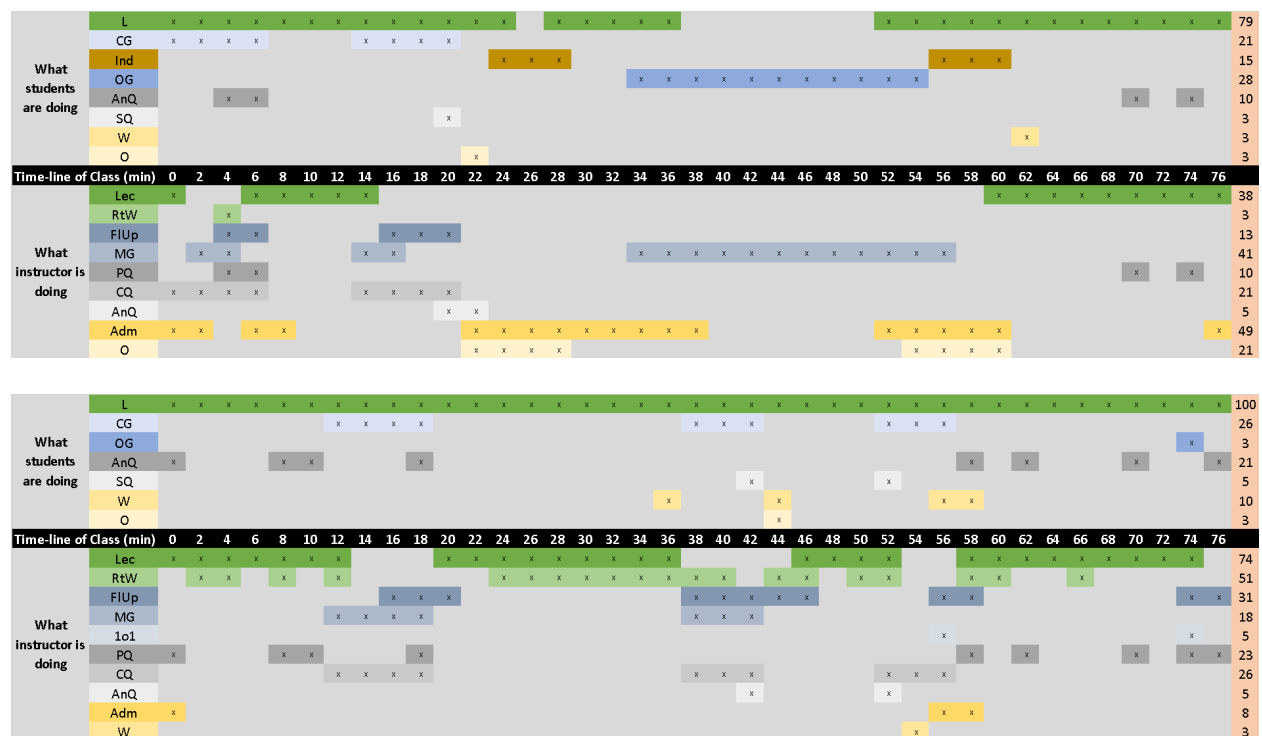


Figure S3. COPUS code timelines for Course Three. Observation day one on top and day two on bottom. Code abbreviations and colors are provided in Table S1. Values at the end of each row are the percentage of 2-minute time blocks the code was observed.

In Course Three, day one included two topic cycles around intermolecular forces (heats of vaporization/fusion and viscosity/surface tension/capillary action), day two included three topic cycles around phase changes (vapor pressure, heating curves, and phase diagrams). The instructor utilized prepared slides when introducing a topic, which included figures and graphs/plots from the text. These slides were annotated as needed. The examples presented to students during each cycle were calculation-based and, as indicated by the instructor, were either

straight from or slight modifications (i.e., different numeric values) to those presented in the pre-class videos. The day one examples were one-step processes (e.g., energy change given an amount and standard heat value), some on day two included more multi-step calculations (e.g., energy change along multiple sections of a heating curve). For the day one non-typical activity, each group of three students were given a set of cards. Each card displayed a molecular structure and the molecule's boiling point. One student (the arbiter) selected two cards and formulated an intermolecular force question about them (e.g., Which has stronger forces?). The other two students raced to identify which molecule answered the question.

Course Four

Class sessions revolved around small-group completion of instructor-authored problem sets. Students were observed working in their assigned groups (WG, Figure S4) during a majority of the time-blocks (79% of each day's two 2-minute time blocks). During these times the instructor and learning assistants were observed moving through the class guiding each group's work (MG, 74% and 68%) and responding to their questions (1o1, 77% and 68%). During the last ~20 minutes of class, the instructor concluded the activity by posing whole-class questions (PQ, 21% and 16%) on the focal points of the day's content. During this time, the instructor was observed providing clarifying information by lecturing (L, 23% and 24%), working a sample problem (RtW, 21% and 16%), and/or through follow-up questions (FIUp, 23% and 24%). During these wrap-up sessions, students were observed listening to the instructor (L, 23% and 29%) and answering questions (AnQ, 21% and 16%). These timelines matched a cluster 6 COPUS Profile indicating a student-centered instructional style.

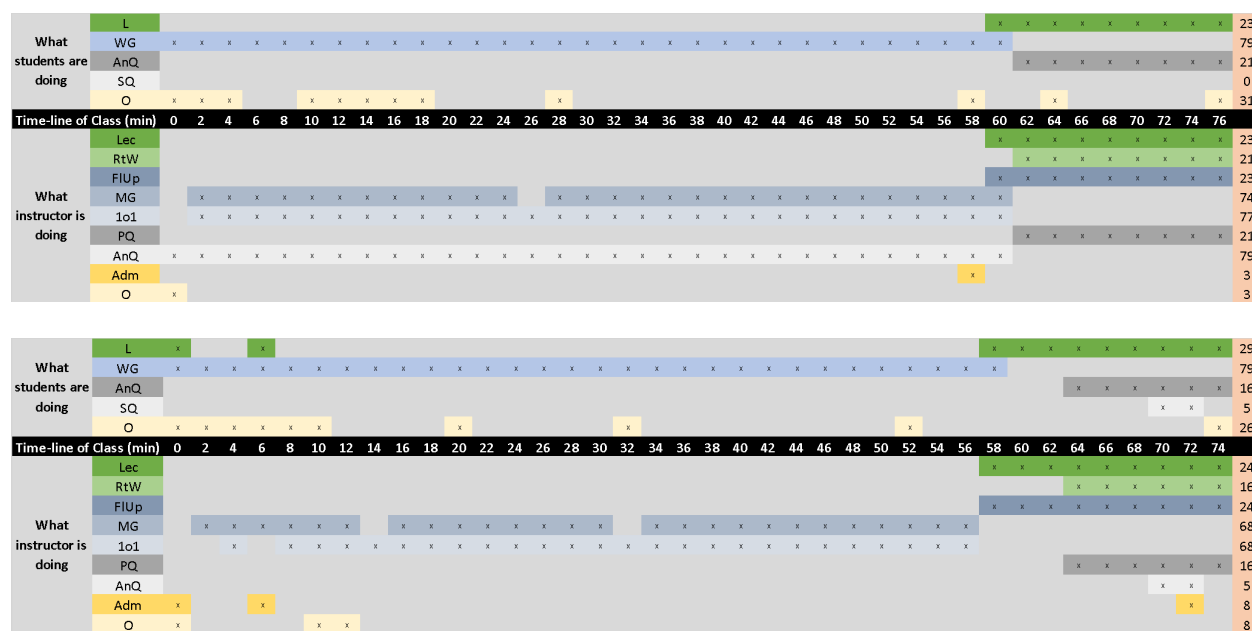


Figure S4. COPUS code timelines for Course Four. Observation day one on top and day two on bottom. Code abbreviations and colors are provided in Table S1. Values at the end of each row are the percentage of 2-minute time blocks the code was observed.

The problem sets in Course Four contained little to no guiding information and focused on a single topic from the assigned pre-class videos. The formal charge problems on day one contained a series of items asking students to first describe/explain the concept, then requiring

them to apply it to a series of structures, and finally use the concept to explain the difference in atom connectivity for given formulas. On day two, the problems involved a series of molecular and ionic formulas for which students were asked to determine the electron pair and molecular geometries as well as to predict their bond angles.

Course Five

Class sessions revolved around small-group completion of an online instructor-authored worksheet. Students were observed working in their assigned groups (WG, Figure S5) during almost all time-blocks (92% and 98% of each day's two 2-minute time blocks) and entering responses on a laptop. The instructor was observed to be consistently moving through the class guiding each group's work (MG, 75% and 85%) and responding to their questions (1o1, 86% and 98%); no learning assistants were present. Intermittently throughout the class time, the instructor was observed posing clicker questions (CQ, 14% and 15%); students were observed working on them individually (Ind, 14% and 15%) and presented their answer using colored response cards. The instructor was rarely observed presenting information (Lec, 0% and 2% and FIUp, 11% and 10%), with occurrences correlated to the clicker questions. Students were observed listening to the instructor (L, 42% and 27%) mainly when administrative details (Adm, 36% and 27%) were given or when check-in questions were being delivered. The COPUS Profile for these timelines matched cluster 5, which represents a student-centered instructional style.

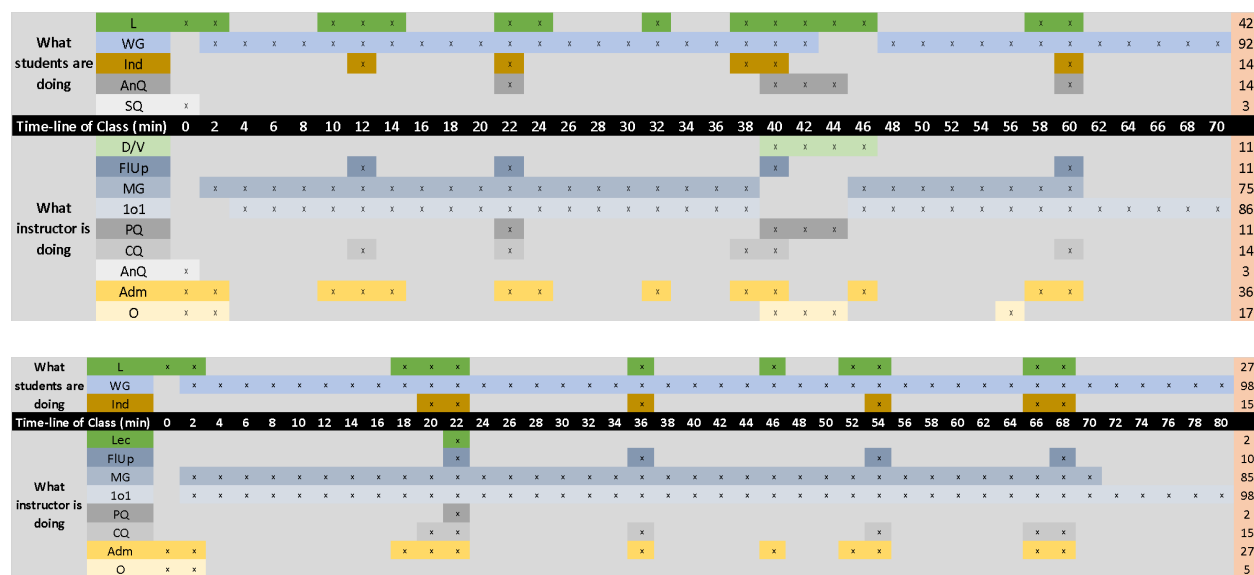


Figure S5. COPUS code timelines for Course Five. Observation day one on top and day two on bottom. Code abbreviations and colors are provided in Table S1. Values at the end of each row are the percentage of 2-minute time blocks the code was observed.

The online workbook used in Course Five contained POGIL-style worksheets. Each contained a number of models with a figure and limited text introducing the context of the figure. The class only met once per week and, as noted by the instructor, the worksheets were aligned with the more foundational or difficult topics covered in the video lectures assigned that week. The worksheets from the observed days covered the topics of electron configurations and periodic trends on day one and bonding on day two. Each model was followed by a series of questions. Initial questions asked students to extract and explain information from the model (i.e., key

questions in the POGIL framework)⁵⁸ to acquire new knowledge. These were followed by exercises requiring students to use the information presented in the model.

Pre-Class Material (PCM) Survey Development

The survey was developed through an iterative process that included two rounds of focus groups and one round of a pilot survey conducted in previous semesters of the courses. Brief descriptions of these processes are outlined in this section. The pilot version of the items can be found in supplemental Table S4. Final item versions are embedded throughout the manuscript and supplemental text (in bold) as well within manuscript Tables 5-7 and Supplemental Tables S5-S9.

Focus Group Participants

During a developmental phase of the project, students were recruited from Courses Two, Four, and Five. An announcement regarding the focus groups was placed at the end of a pre-semester survey administered as part of the larger project. Students noted their interest in participating in a discussion group about their flipped course and provided contact information for scheduling. Prior to an on-campus visit, all students who expressed interest were notified of the focus group times and asked to respond indicating which group(s) fit their availability. Participating students were compensated with a \$10 gift card.

Focus Group Data Collection and Analysis Protocols

Focus groups were conducted in person. Due to timing and other constraints, groups were not conducted with students from Courses One and Three. As researchers from the project were on each campus for a minimum of three days, a range of days and times were provided for the focus groups. For ease of access, each group met in a private location in or near the building where the flipped chemistry course was taught. Each group was conducted by two researchers. The first interviewer initiated the focus group questions, the follow-up questions, and managed discussions among participants. The second interviewer took notes on the discussion and provided additional follow-up questions or asked clarifying statements as needed. All focus groups were video recorded and these recordings were coded by one or two researchers.

The first round of focus groups were conducted with students from Course Four and were general in nature, asking about overall study habits with regard to the flipped course (Table S2). Two coders conducted reviews of the video recordings from each group to develop a codebook and subsequently code each groups' responses. Independently, each coder reviewed three videos and documented the salient responses for each question, creating their own preliminary codebook. The coders then met to discuss their codebooks and generated a single codebook. The codebook was then independently used by each coder to review the remaining videos. This coding process produced Cohen's kappa scores >0.81 for each video, indicating near perfect inter-rater reliability¹. The coded responses from these groups provided insights for the development of more formulated questions around students' use of the video resources for the pilot survey and an additional round of focus groups.

¹ Cohen, J., A coefficient of agreement for nominal scales. *Educational and psychological measurement* **1960**, *20* (1), 37-46.

Table S2. Pre-class preparation focus group questions.

As your instructor requests that you come to class prepared to engage in learning activities, the next set of questions is about your pre-class preparation.
1. On average, how much time do you spend preparing for class?
2. What do you typically do to prepare for class?
3. Are there other things you could/should be doing prior to class?
4. What are the limitations to doing these additional things?

The second round of focus groups, conducted with students in Courses Two and Five, focused on questions specific to the pre-class videos used in the flipped courses (Table S3). A primary coder reviewed two focus group videos, one from each institution, and created an initial codebook. A secondary coder used the initial codebook to independently code the same two videos, following which they met to discuss discrepancies and make codebook modifications. Using the modified codebook, each reviewer independently coded two additional videos. This coding process produced Cohen's kappa scores >0.81 for each video, indicating near perfect inter-rater reliability¹. All subsequent videos were coded by the primary coder using the modified codebook. Responses from these groups were used when developing the final version of the survey from the pilot version.

Table S3. Pre-class preparation focus group questions.

1. Do you regularly watch the posted videos?
<i>Follow-ups: For those that do</i>
<ul style="list-style-type: none">• How do you watch them? (e.g., from start-to-finish completely, skip around, fast forward through, etc.)• What do you do when watching them? (e.g., nothing but watch and listen, take notes, work on quiz or homework questions, etc.)• When do you typically watch them for the first time? (e.g., evening before class, end of week, etc.)• Do you ever re-watch them? If so, when?
<i>Follow-ups: For those that do not</i>
<ul style="list-style-type: none">• Have you ever accessed them?• If yes, why do you not regularly watch them?• If no, why do you not access them?
2. Do you find the videos helpful for learning the material? Please explain why or why not.
3. What do you like or dislike about the videos? Please explain.

Focus Group Results

A total of 56 students participated in focus groups, 10 from Course Two (7 groups), 24 from Course Four (7 groups), and 22 from Course Five (8 groups). Groups varied in size from a maximum of 4 to a minimum of 1. Due to scheduling issues and no-shows, some planned focus group meetings ended up including only one individual. While not ideal, we valued each student's time and input and therefore did not cancel smaller groups nor reject individual input.

From the first round of focus groups, conducted with students from Course Four, some general response themes emerged; such as, that students were watching the videos to prepare for class (79%) as well as re-watching the videos later (71%). Additionally, students reported general aspects that they liked about the videos (67%), such as being able to watch at their own pace, and also disliked (88%), such as not being able to ask questions. These general response themes were used to create specific questions around students' usage and perceptions of the pre-class videos, which were administered during the second round of focus groups and on the pilot survey.

The second round of focus groups, conducted with students from Courses Two and Five, provided similar and additional insights to students' use and perceptions of the pre-class videos. For example, with regard to how students watched the videos, only 37% reported that they watch straight through, with many reporting that they pause/rewind (87%), watch at a different pace (28%), or skip/fast-forward through (37%). Students expressed many positive perceptions of the videos such as being able to watch at their own pace (64%) or whenever and/or wherever they want (45%), and commented positively about the structure (55%) and length (48%) of the videos. Their negative perceptions focused around not being able to ask questions (33%), that the videos did not keep their attention (21%), or that they did not feel there were enough problems or explanations provided (33%). In all, the responses from the two rounds of focus groups were informative in providing clarity to some of the pilot survey responses and to adjust the exact wording of items and responses for the final survey.

Pilot Survey Population and Administration

Students were recruited to participate in the pilot survey from Courses One, Two, and Five. Due to timing and other constraints, pilot survey data was not conducted in Courses Three and Four. Survey deployment in each course was coordinated to take place midway through the course during a non-exam week. The instructor was provided a brief script to make an initial in-class announcement regarding the survey. A note similar to the script was posted on the classroom management platform of each course. Students who were interested in participating clicked on a link to the Qualtrics survey that was part of the announcement note. Some instructors offered a nominal amount of extra-credit points for accessing the survey.

Pilot Survey Data Collection and Analysis Protocols

The pilot survey focused on questions related to the pre-class videos and contained a variety of item types including single-, multi-, and open-response formats (Table S4). The survey flow contained logic steps that populated questions, and their associated follow-ups, based on a participant's prior responses. Therefore, the number of participants was not constant across items and not every individual was presented with each question or follow-up.

Response percentages for single- and multi-response item types were calculated based on the number of participants who were presented with the question. Open-response items were coded and response percentages per code were calculated based on the number of participants who were presented with the question. A primary coder reviewed an aggregated dataset, that contained the item-by-item responses from each institution, and created a codebook for each item. A secondary coder used the codebook to independently code all responses across items, following which they met to discuss any discrepancies. The coders discussed any noted discrepancies and came to consensus on the codes and their use.

Table S4. Pre-class preparation pilot survey questions.

<p>1. Do you regularly watch the posted videos? (Yes/No Radio Buttons)</p> <p><i>(If Yes)</i></p> <ul style="list-style-type: none"> ● Please explain how you watch them. (e.g., from start-to-finish completely, skip around, fast forward through, etc.) (Open-response) ● What do you do when watching them? (e.g., nothing but watch and listen, take notes, work on quiz or homework questions, etc.) (Open-response) ● When do you typically watch them for the first time? (Single-response question) <ul style="list-style-type: none"> ○ The evening before class ○ Within a few hours before class ○ Within a few hours after class ○ At the end of a week of class ○ Other (Please explain) (Open-response) ● Do you ever re-watch them? (Yes/No Radio Buttons) <p><i>(If Yes)</i> When do you re-watch the videos? (Multi-response)</p> <ul style="list-style-type: none"> ○ At the end of a week of class ○ When studying for a quiz or exam ○ When working on the homework ○ Other (Please explain) (Open-response) <p><i>(If No)</i> Why do you not re-watch the videos? (Multi-response)</p> <ul style="list-style-type: none"> ○ I don't find them useful to watch more than once ○ I watch other videos to get a different perspective than the ones posted ○ They typically don't match well to the quiz or exam material ○ Other (Please explain) (Open-response)
<p><i>(If No)</i></p> <ul style="list-style-type: none"> ● Have you ever accessed the posted videos? (Yes/No Radio Buttons) <p><i>(If Yes)</i> Why do you not regularly watch them? (Please explain) (Open-response)</p> <p><i>(If No)</i> Why have you not accessed them? (Please explain) (Open-response)</p>
<p>2. ^aDo you find the videos helpful for learning the course material? (Yes/No/Sometimes Radio Buttons)</p> <p><i>(If Yes)</i></p> <ul style="list-style-type: none"> ● Please explain why you find the videos helpful for learning (Open-response) <p><i>(If No)</i></p> <ul style="list-style-type: none"> ● Please explain why you do not find the videos helpful for learning (Open-response) <p><i>(If Sometimes)</i></p> <ul style="list-style-type: none"> ● Please explain why you only sometimes find the videos helpful for learning (Open-response)
<p>3. ^aWhat do you like about the videos? (Open-response)</p>
<p>4. ^aWhat do you dislike about the videos? (Open-response)</p>

^aItems 2-4 presented only if response to item 1 = yes

Pilot Survey Results and Survey Refinement

When presenting the results from the pilot survey, any resulting refinements for the final survey will also be noted. The pilot survey resulted in 263 responses from Course One, 27 from Course Two, and 22 from Course Five.

The first question, “**Do you regularly watch the posted videos?**”, resulted in 94 (38%), 24 (89%), and 15 (68%) ‘Yes’ responses from each institution respectively. Students who responded with ‘No’ were not asked any of the additional follow-up questions about how they interacted with the videos. When this same question was asked of students in the second round of focus groups it was discovered that many students who did not *regularly* watch the posted videos still watched the videos occasionally or when they were confused about a specific topic. Therefore, the wording of this item was updated for the final survey to “**How many of the assigned videos have you watched?**” with options of ‘All’, ‘Most’, ‘Some’, or ‘None’. All

students who selected ‘Some’, ‘Most’, or ‘All’ would be directed to answer the follow-up items about how they interacted with the videos. Students who responded ‘None’ would be directed to a follow-up item asking if they have ever accessed them, with appropriate follow-ups based on their response.

The 133 participants who responded that they did regularly watch the videos were presented with the first follow-up question “**How do you watch them?**”. Open-ended responses to this question generated several response categories: From start to finish (43%), pausing/rewinding (13%), playing at faster speed (4%), skipping around (15%), watch multiple times (6%), take notes (13%), and work problems (3%). For the responses (n=133) from the second follow-up question “**What do you do when watching them?**”, the majority of students responded that they take notes (77%), while fewer reported just focusing on the video (32%) or working practice problems (25%). The response categories for these two follow-up questions were similar to those from the focus groups, therefore, each category was retained and multi-response options were generated for the final survey version.

The next question focused on when students typically watched the videos for the first time (as opposed to re-watching), with options worded around the timing of the class itself. While students at each institution used many of the provided categories, the ‘Other’ option was selected quite frequently, 15% each for Courses One and Two and 60% for Course Five students. Within the textbox provided, students’ explanations of when they watched were typically based on aspects such as when topics/material was being covered in class or around doing homework/studying for exams. These sentiments were also noted in the second round of focus group responses. Therefore, the response options for the final survey were modified to reflect these types of timings for viewing.

Next, the pilot survey included the ‘Yes’/‘No’ item, “**Do you ever re-watch [the videos]?**”. Of the 143 responses, 64% noted that they did re-watch the videos. These students were administered the follow-up question “**When do you typically re-watch them?**”, reporting that they re-watched when studying (77%), doing homework (46%), or at the end of the week (13%). The 52 students who reported that they did not re-watch were asked why they did not and reported that they watched other videos instead (42%), found the videos were not useful the first time, or when selecting the ‘Other’ response (31%), wrote that they relied on the notes they wrote when watching the videos for first time. When the initial ‘Yes’/‘No’ item was asked to students in the second-round focus groups, such as that they had not re-watched an entire video or that they would re-watch portions of the videos that they found confusing or contained a topic on which they needed clarification. Therefore, this item was modified in the final survey to read, “**Have you ever watched a video (or part of a video) more than once?**”

In addition to the items related to how the students interacted with the videos, the last items on the pilot survey and in the focus groups were about students’ perceptions of the videos. Students were asked “**Do you find the videos helpful for learning the course material?**” and what they liked and/or disliked about the videos. While the responses to these open-ended items were coded separately, their resulting response categories were similar. The categories of what students found helpful about the videos (i.e., ability to pace watching, easy to understand, reinforce material) map directly onto what students reported that they liked about the videos.

Similar overlaps were found between why they reported that the videos were not helpful (i.e., not engaging/can't focus, too long or no time to watch, prefer other methods) and their reported dislikes. In the focus groups, students' likes and dislikes were expressed when responding to the questions about what was helpful and not helpful, often no additional insights were provided when they were asked about likes and dislikes. These outcomes revealed that both sets of items were not needed on the final survey. Students' reported likes/dislikes about the videos were combined with the themes that arose from what students found helpful/not helpful about the videos when generating response categories for the final survey items.

Supplementary Tables

Table S5. Response percentages, by course, to survey item “Why have you not watched all of them [assigned videos]?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	57	29	117	109	8
^aWhy have you not watched all of them? (Select all that apply)					
Response categories and options	Percentage of students selecting an option				
<i>General excuses</i>					
Not enough time	53	45	48	32	50
I forget to watch them sometimes	47	45	32	39	25
They are too long***	16	7	15	35 ^b	25
<i>Not helpful</i>					
They do not help my learning***	2	17	9	27	13
I only watch them when I need clarification on course material***	75 ^c	34	39	29	38
<i>I prefer other...</i>					
videos than the ones my instructor posts	7	7	5	14	13
learning resources	26	24	23	35	13

^aSurvey item presented if ‘Most’ or ‘Some’ response options were selected for initial question, shown in Figure 2. *** $p < 0.001$. ^bSignificant pairwise comparisons ($p < 0.05$) between Course Four and Courses Two ($w = 0.25$) and Three ($w = 0.23$). ^c Significant pairwise comparisons ($p < 0.05$) between Course One and Courses Two, Three, and Four.

Table S6. Response percentages, by course, to survey item “Why have you never re-watched the videos?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	6	7	40	44	16
^aWhy have you never re-watched the videos? (Select all that apply)					
Response options	Percentage of student responses to item				
I refer to the notes I take the first time I watch**	33	57	70	59	100
I watch other videos to get a different perspective than the ones posted***	67	29	3	23	6
I typically understand the material after watching just once	33	14	30	36	56
They typically don’t match well to the quiz or exam material	0	29	13	20	0

^aSurvey item presented to everyone who selected ‘No’ to question in Figure 3. ** $p < 0.01$. *** $p < 0.001$.

Table S7. Response percentages, by course, to survey item “When do you re-watch the videos?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	58	26	193	232	42
^aWhen do you re-watch the videos? (Select all that apply)					
Response categories and options	Percentage of student responses to item				
<i>I re-watch parts of a video when...</i>					
...I have missed something the first time	78	65	78	72	76
...I need clarification at a later time (e.g., for homework or when completing a lab)*	88	69	81	72	79
...studying for an exam*	43	35	61	53	50
<i>I re-watch an entire video...</i>					
...to supplement my notes**	16	12	24	26	2
...when studying for an exam	28	15	28	32	14

^aSurvey item presented to everyone who selected ‘Yes’ to question in Figure 3. *p < 0.05. **p < 0.01.

Table S8. Response percentages, by course, to survey item “When you watch the videos, how do you watch them?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	64	32	233	276	58
^aWhen you watch the videos; how do you watch them? (Select all that apply)					
Response categories and options	Percentage of student responses to item				
<i>Pacing of viewing</i>					
From start-to-finish at normal speed**	66	59	58	49	34
From start-to-finish at varying (faster or slower) speed***	23	16	24	33	69
I pause and/or rewind while watching	73	63	61	64	52
I skip around or fast-forward through sections	23	16	27	20	19
<i>Blocking of viewing</i>					
I watch the assigned videos in one sitting***	25	13	32	55 ^b	60 ^b
I spread out watching the assigned videos throughout the day or week***	20	19	25	6	28

^aSurvey item presented to everyone except those who selected ‘None’ for initial question, shown in Figure 2. ** p < 0.01. *** p < 0.001. ^bSignificant pairwise comparisons (p < 0.05) between Courses Four and Five and Courses One (w = 0.24 and w = 0.36, respectively), Two (w = 0.26 and w = 0.46, respectively), and Three (w = 0.23 and w = 0.23, respectively).

Table S9. Response percentages, by course, to survey items “Were the videos NOT helpful to your learning? If so, in which ways were they NOT helpful?”

	Course One	Course Two	Course Three	Course Four	Course Five
Students, n	42	25	173	254	46
^aWere the videos NOT helpful to your learning? If so, in which ways were they NOT helpful? (Select all that apply)					
Response categories and options	Percentage of student responses to item				
<i>Do not meet learning expectations</i>					
I am unable to ask questions or interact with the instructor***	40	40	28	62	59
They do not contain enough practice problems**	26	36	39	51	41
<i>Not relevant to course</i>					
They are too basic	14	20	21	26	13
The explanations are too difficult	7	16	10	12	2
They have a different focus than the class materials***	12	28	10	37	11
They contradict the class material***	0	0	1	9	4
<i>Don't hold attention</i>					
They are too long***	33	8	25	27	57
They are boring or not engaging***	17	12	29	40	17
The material presented is redundant	19	0	13	11	9
<i>Poor quality/disorganized</i>					
They are low quality making it difficult to see and/or hear***	5	0	0	27	9
They are confusing and/or disorganized**	0	4	7	14	4

^aSurvey items presented to everyone except those who selected ‘None’ for initial question, shown in Figure 2. *p < 0.05. **p < 0.01. ***p < 0.001.