Effects of classroom "flipping" on content mastery and student confidence in an introductory physical geology course

Jason P. Jones, David A. McConnell, Jennifer L. Wiggen & John Bedward

View supplementary material ↗	Published online: 01 Apr 2019.
Submit your article to this journal ↗	Article views: 456
View related articles ↗	View Crossmark data ↗
Citing articles: 5 View citing articles ↗	

Routledge
Taylor & Francis Group

Check for updates

Effects of classroom "flipping" on content mastery and student confidence in an introductory physical geology course

Jason P. Jones[a] , David A. McConnell[a] , Jennifer L. Wiggen[a] and John Bedward[b]

[a]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina 27695; [b]Department of Science, Technology, Engineering, and Mathematics, Buena Vista University, Storm Lake, Iowa 50588

ABSTRACT

Incorporating active learning strategies into introductory STEM courses has been shown to improve student outcomes, however, these activities take class time to execute. The question of how to implement these effective strategies without sacrificing a significant volume of content coverage has led to the development of a "flipped" model of instruction. This flipped model requires students to take responsibility for learning some basic concepts prior to attending class so the instructor can use newly freed class time to incorporate active learning activities. This study investigated the impact of implementing a partially flipped class format on student exam performance and confidence across four semesters of a large-enrollment physical geology course. Basic geology content was presented as pre-class homework assignments using short instructional videos (*Geoscience Videos*) that were created following empirically-derived methods of effective multimedia design. The videos facilitated an increase in the proportion of content that could be communicated outside of class and allowed for an augmentation of in-class activities on more complex geology concepts. We compared student performance and confidence across semesters and found; (a) students were able to learn the basic content as effectively as they had when it was presented in class; (b) students improved their performance on some content during summative exams; and, (c) student confidence significantly varied on some topics as a result of the course alterations. As a result, we posit that the flipped model can provide valuable opportunities to increase student learning as long as students are supported via out-of-class homework and feedback on their level of understanding regarding topics they are learning prior to attending course meetings.

Introduction

What determines the level of student learning in introductory geoscience courses? Is it what happens in the classroom? Or is it what occurs when students consider course content outside of class? This study sought to examine the balance between these two aspects of the learning process within a face-to-face university introductory science course. Specifically, we examine pre-class preparation where students gather information regarding foundational course content and student interaction with instruction during class meetings.

The traditional approach to providing students with the requisite background knowledge for subsequent class meetings has typically been in the form of course reading assignments (Burchfield & Sappington, 2000). Unfortunately, many students do not read the textbook as preparation for the class meetings

(Burchfield & Sappington, 2000). Podolefsky and Finkelstein (2006) found that only 13% of students in introductory physics classes reported that they "often" (>80% of the time) read the book before class. In contrast, most students appear to consider their textbook as a reference source to support preparation for exams rather than as a resource to be used as a primer for in-class learning (Podolefsky & Finkelstein, 2006). This is exacerbated by evidence that most introductory geoscience textbooks are not written in a pedagogically-effective way, utilizing an overabundance of technical terms and dissociating text from pertinent figures (Kortz, Grenga, & Smay, 2017). Students may be inadvertently encouraged to ignore pre-class assignments if instructors subsequently present many of the same concepts and terms in class (Crouch & Mazur, 2001). The result is an inefficient

---

instructional system where the parameters of student learning are defined almost exclusively by what occurs inside the classroom.

Discipline-based education research has revealed that we can enhance student learning in STEM disciplines through the application of active learning pedagogies (e.g., Singer, Nielsen, & Schweingruber, 2012; Freeman et al., 2014). In a general sense, the concept of "active learning" is derived from the concept that students are active participants in their own learning process, with Bonwell and Eison (1991) originally defining the concept as "instructional activities involving students doing things and thinking about what they are doing" (p.5). In the classroom environment, active learning places the emphasis of instruction on students and their experiences as opposed to having them be passive recipients of information (Mazur, 1997). For the educator, active learning involves providing students some combination of the following elements during in-class instruction: a) information and ideas (i.e., the content); b) experiences that allow them to either complete a task or observe some phenomena related to the content; and c) opportunities for them to reflect on their learning as individuals or by discussing with peers (Fink, 2003; see also, McConnell et al., 2017). A meta-analysis comparing active learning to more traditional didactic models reported average student learning gains of approximately 0.5 standard deviations (∼6%) and a 1.5x reduction of the student drop, fail or withdraw (DFW) rate (Freeman et al., 2014).

Despite the measured benefit of active learning, the challenge for instructors is how to find time to incorporate these strategies into their courses without sacrificing content coverage (Crouch & Mazur, 2001). This dilemma contributed to the development of the "inverted" or "flipped" learning model (Lage, Platt, & Treglia, 2000; Bishop & Verleger, 2013; Gross, Pietri, Anderson, Moyano-Camihort, & Graham, 2015). Flipped courses place some of the responsibility for learning basic content onto the students outside of the class and is often facilitated by the use of online assessments, often paired with quizzes (e.g., Lage et al., 2000; Gross et al., 2015). These required assessments ensure that students have a stake in interacting with the basic content prior to attending class. Additionally, by moving some content outside of the classroom experience, this approach can provide additional time during class meetings for instructors to incorporate activities that allow students to tackle more demanding or application-based tasks in the collaboration with their peers (Evans, 2011; Strayer,

2012; Tucker, 2012; Gajjar, 2013). Additionally, the practice of "flipping" may allow the instructor to go into more depth on a topic and to present students with opportunities for formative assessments so that they can practice applying a new concept and assess their comprehension (Freeman, Haak, & Wenderoth, 2011).

Approaches to classroom flipping can vary in scope and relative volume of materials that are moved outside of the classroom. The level of modification and volume of material moved can differ by instructor goals or course level (e.g., higher-level courses potentially giving more responsibility to students). In this example, which we characterize as a "partial flip", we moved some portions of instruction to pre-class assignments (approx. 20% of content), with class meetings representing a mix of content delivery and active learning strategies. This is opposed to a "full flip" which moves all delivery of content outside of the classroom and leaves course meetings to exclusively focus upon applications of concepts and in-class activities.

Investigations into the effectiveness of flipped environments is subject to wide variability in both study designs and reported results. A common experimental design investigating flipped classrooms compares a newly-adopted flipped approach to the course (including pervasive active learning elements) against a control semester that followed the traditional didactic approach, minus the active learning activities (e.g., Heyborne & Perrett, 2016; Schultz, Duffield, Rasmussen, & Wageman, 2014; Strayer, 2012; Tucker, 2012; Tune, Sturek, & Basile, 2013). Given the independent effects of active learning on improving student performance in STEM courses (e.g., Freeman et al., 2014), it is apparent that observed benefits cannot be solely attributed to the flipped course design. Other studies have attempted to control variables via the comparison of courses utilizing active learning and those using active learning with flipped components.

One such study, Gross et al. (2015), compared five years of data from an upper-level biochemistry course taught with varied methods of content delivery (three iterations of traditional instruction and two years of flipped instruction with active learning activities). In the flipped semesters, researchers recorded video lectures akin to their prior face-to-face lecturing, broke them into 5-20 minute segments, and made viewing them optional prior to course meetings. Instructors used this newfound class time for active problem-solving and team-based learning activities. Researchers found significant performance gains for students in the revised version of the course in comparison with students from prior semesters taught in a traditional

format (Gross, Pietri, Anderson, Moyano-Camihort, & Graham, 2015). Gains on exam scores were greatest for female students and for students entering the course with lower GPAs (Gross et al., 2015).

Another study that sought to control variables related the adoption of flipping, Jensen et al. (2015), compared two sections of an undergraduate introductory biology course that both utilized an active learning approach, but one section was taught in a flipped format and the other without. In the study, students in the flipped condition were required to complete quizzes based on a mixture of background readings, demonstration videos, and probing questions that were delivered in the online course management system (CMS) prior to each class meeting (3 per week; 39 in total) (Jensen, Kummer, & Godoy, 2015). Students then applied their knowledge of the content in novel situations via in-class group activities. In contrast, students in the non-flipped approach completed the same exploratory activities in-person via group work, but finished the lesson via a required CMS quiz that was completed after class and consisted of the same information contained in the flipped condition. Results found no significant differences between sections varying only in pre-class consumption of background material (Jensen et al., 2015). These results may be interpreted to suggest that the primary benefit for flipped learning may be found within the active learning approaches facilitated in the classroom by its adoption.

Student confidence

While performance (measured by summative exams) and persistence (measured by the DFW rate) are benchmarks in determining student success in academic environments, further insight into how students are interacting with course activities can be learned from the collection of student confidence measures. Bandura (1997) represented the variability of the student learning process as the result of personal, environmental, and behavioral factors. These factors, when considered in practice, lead to the construct of self-regulated learning (SRL) and several studies (for example, see Zimmerman, 1990) sought to apply these principles to classroom learning environments (Schraw, Crippen, & Hartley, 2006).

Self-regulated learning is typically divided into three primary subcomponents (see Panadero, 2017, for a review): metacognitive awareness (e.g., awareness of one's own skills and abilities), behavior (e.g. effective isolation and use of strategies), and motivational

control (Schraw et al., 2006; Zimmerman, 2008). These skills are put into practice by self-regulated learners via a three-phase process employed during a learning task: planning (e.g., "what study strategies will I use to prepare for this exam?"), monitoring of performance (e.g., "do I know the answer to this question?") and reflection (e.g., "did I meet my goals?"). These reflections on relative success of the learning process are then recycled into the planning phase for the next learning task. Student confidence during learning tasks is housed within the second performance phase of the SRL feedback loop. The confidence judgment a student makes is determined by a combination of the student's prior knowledge of the topic, characteristics of the text, characteristics of the question, and guessing (Dinsmore & Parkinson, 2013). Specifically, the measure of confidence that captures the level of student belief that they selected the correct answer while responding to an exam question represents a *judgement of learning* (Schraw, 2009). As this study involves removing and augmenting instruction from in-class meetings, we sought to collect student confidence related to each exam question via judgments of learning to potentially signal any affective response to the course changes aside from overall exam performance.

Videos as a learning element

One may ask, is all flipping created equal? One version of a flipped class format introduces some course content prior to lecture using instructional videos and related assignments (Moravec, Williams, Aguilar-Roca, & O'Dowd, 2010). Students can also use these videos for content review and they may prove valuable for learners of different abilities as students are able to pause, repeat, and re-watch the videos as necessary (Schultz et al., 2014). The use of video has been shown to increase students' attention and engagement (e.g. Green, et al., 2003; Jha, Widdowson, & Duffy, 2002; Zhang, Zhou, Briggs, & Nunamaker, 2006), increase motivation and self-efficacy (Bennett & Glover, 2008) and improve understanding (e.g. Choi & Johnson, 2005; Seeling & Reisslein, 2005; Zhang et al., 2006). Additionally, access to video lessons in flipped classes can provide students with autonomy and control in the study process and may avoid expertise reversal effects as students with more conceptual understanding can skip over scaffolds intended for more novice learners (Kalyuga, Ayres, Chandler, & Sweller, 2003). Stelzer et al. (2009) described how the introduction of video-based pre-class multimedia modules (MMs) in introductory physics classes resulted in an improvement of student performance on an assessment immediately

after completion of the module, and again on a test two weeks later. Performance of a group of students who were exposed to the MMs was approximately one grade level better than similar student populations that reviewed equivalent information in textbook format (Stelzer, Gladding, Mestre, & Brookes, 2009). The MMs consisted of a series of brief "acts" combined in a narrated 15-minute-long video with some embedded multiple choice questions. Further, when student results from questions in the MM-supported classes were compared to those from prior years that relied on textbook reading alone, the MM-supported students scored higher (average 57% correct answers vs. 49%; Chen, Stelzer, & Gladding, 2010). The degree of improvement was not consistent among concepts, suggesting that some MMs were more effective than others (Chen et al., 2010).

Videos employ visual, audio, temporal, and other representational symbol systems to transmit information that may be difficult to convey through a textbook or lecture format (Tantrarungoroj, 2008). Some of this information may be sequences in motion (e.g., movement along a fault or how Coriolis force modifies wind direction as a function of latitude), or perspectives that are difficult to observe in real life, whether due to their temporal, spatial, or remote nature (e.g., diving under the Antarctic ice or motions of tectonic plates and their effects; Wetzel, Radtke, & Stern, 1994). Towards this end, the work of Mayer in aligning the characteristics of students' cognitive abilities with multimedia design has generated three primary assumptions of multimedia learning (see Mayer, 2002, for a review): 1) Students possess two separate channels for processing audio and visual information (Baddeley, 1999); 2) Presenting information efficiently via both audio and visual channels can increase the net amount of information processed (in effect increasing students' working memory capacity; Chandler & Sweller, 1991; Mayer, 2002); and 3) Learners are "active processors who seek to make sense of multimedia presentations" and not passive "tape recorders" of presented information (Mayer, 2002, p. 36). Multimedia that is designed with these assumptions in mind have generated four effects that have been shown to increase learning in students consuming multimedia content (Mayer, 2003). They are: 1) multimedia effect - displaying words and pictures simultaneously is more effective than words alone; 2) coherence effect - eliminating extraneous details in media improves learning; 3) spatial contiguity effect - images and descriptive words are more effective when in close proximity; and, 4) personalization effect - more conversational language is more

effective than technical or formal language often featured in textbooks (Mayer, 2003). Though these elements of multimedia design have been shown to be effective both on paper (e.g., graphics, posters) and in videos, these effects have the greatest potential when paired with the added expansions of cognitive processing ability that video provides (Mayer, 2002).

To utilize the benefits of effective multimedia design and apply them to the introductory geoscience classroom, we began to create videos in attempts to follow the principles of effective multimedia design (Dixon & McConnell, 2014). These Geoscience Videos are short, content-related videos designed to convey foundational concepts of introductory geoscience courses. The videos are typically 5-7 minutes long and follow a standard format consisting of a mix of learning objectives, basic content topics, brief text coupled with images (e.g., maps, diagrams, models, geologic features) and/or embedded video clips (e.g., demonstrations), a formative assessment, a summary reflection activity, and conversational narration throughout (which is also available as closed captions). In addition to being embedded in the required course homework assignments, videos were also available publically on a YouTube channel (www.youtube.com/c/geosciencevideos; Dixon, McConnell, & Bedward, 2015; Dixon, 2016).

We sought to investigate the effects on student exam performance and confidence of partially-flipping a university-level introductory physical geology course via the introduction of video-based assignments prior to class meetings. Over four study semesters, course pre-class reading assignments were gradually replaced with video-based assignments. As a result, direct in-class instruction of many introductory concepts (e.g., igneous rock classification, identifying different types of volcanoes) was removed from the course. This change resulted in more time becoming available in class to address challenging concepts, and to utilize additional active learning activities to support learning. Specifically, we sought to determine how the increase in pre-class video content and subsequent changes in in-class instruction across the study semesters affected:

1. Student exam performance and mean level of confidence related to concepts now communicated outside of the classroom.

    -   It is hypothesized that there would be no significant differences in student performance or confidence on summative exam questions based on content that was now delivered outside of class due to the efficiency (both cognitively and procedurally) of multimedia learning.

2. Student exam performance and mean confidence related to concepts that were covered more extensively in class as a result of the flipped format.

- It is hypothesized that students would perform better and have greater confidence on such concepts due to the positive effects of active learning strategies in science courses.

We seek to generalize any findings and make recommendations for practice to instructors who are considering pursuing a flipped model of instruction.

Methods

We compared student performance and confidence measures on exam questions across four consecutive Fall semesters of a section of an introductory physical geology course at a large public research university in the Southeastern US. Efforts were made to control for many situational factors affecting each cohort in an attempt to mitigate confounding variables. Each target class shared the same instructor, was offered during the same academic semester (Fall) and time of day (early afternoon), and measured student performance and confidence via equivalent exam questions. The format (readings vs. video) of course pre-class (homework) assignments and the platform for the delivery of online practice quizzes were altered during the study period, but the breadth of course topics covered in pre-class assignments and during class meetings was essentially equivalent across study semesters.

Participants and setting

Participant cohorts consisted of four convenience samples of undergraduate students enrolled in a class section of the target course and spread across four Fall semesters (2013, 2014, 2015, and 2016). The class met two times a week with each meeting lasting 75 minutes. Students self-selected into the course via enrollment and class sizes ranged from 77 to 94

students. Students varied in academic rank and in age, with the majority of students being either freshman or sophomore (>75% in each semester) and non-STEM majors (~75% in each semester; Table 1). As is historically the case for this course, each semester's sample population contained a male majority (Table 1).

Course alterations

Throughout this study, the target course was taught utilizing an active learning format and each class (except for days with exams) was preceded by a homework assignment known as a learning journal (McCrindle & Christensen, 1995). Scores on the learning journal activities accounted for approximately 20% of the course grade during each semester. Students had access to class resources such as copies of the lecture slides, lists of unit learning objectives, online practice quizzes and practice exams. All four semesters involved a partially flipped class where some material was presented to students through the learning journals prior to each class meeting. What changed over time was the format of the learning journals (readings were replaced with videos), the amount of material covered before class (greater content coverage in videos), the content of the related lectures (removal of introductory content featured in videos), and the characteristics of active learning components in each class (more activities introduced in class as a result of time gained from removing instruction of content covered in videos). This study seeks to try to tease apart the relative effects of these changes.

Pre-class Learning Journal Activities – Students were expected to complete an online homework assignment known as a "learning journal" before attending class each day. These learning journals numbered nineteen in total with eighteen of them being content-based and one focusing on exam preparation. In Fall 2013 students would read assigned pages (approximately four pages) from the textbook to provide them with requisite background knowledge and basic content understanding prior to course

Table 1. Participant gender, academic rank, and declared major.

	Fall 2013 ($n = 93$)		Fall 2014 ($n = 94$)		Fall 2015 ($n = 77$)		Fall 2016 ($n = 91$)	
	count	%	count	%	count	%	count	%
Male	67	71.28	70	74.47	46	59.74	55	60.44
Female	26	27.66	24	25.53	31	40.26	36	39.56
Freshman	63	67.02	52	55.32	51	66.23	47	51.65
Sophomore	14	14.89	22	23.40	16	20.78	25	27.47
Junior	7	7.45	9	9.57	8	10.39	12	13.19
Senior	7	7.45	11	11.70	2	2.60	7	7.69
Other	2	2.13						
STEM	27	28.72	23	24.47	20	25.97	22	24.18
non-STEM	66	70.21	71	75.53	57	74.03	69	75.82

meetings. Students would then answer a number (∼3-5) of content-related questions on the online course management system (CMS; Moodle) to assess their learning. The questions were presented in a combination of formats (multiple choice, true/false, short answer). At the beginning of the corresponding physical meeting of the class, students would typically respond to a few (∼2-4) related conceptual multiple choice questions which they would answer via a classroom response system ("clickers"). The results would then be displayed and discussed during the first few minutes of the class period. This in-class activity served as a review of homework and provided the instructor with an opportunity to identify any lingering misconceptions.

During the Fall 2014 semester, short, content-related videos were added to eight of the eighteen (∼44%) learning journals in the course to replace some of the required reading assignments. The videos made it possible to introduce more content than readings alone could in the equivalent amount of time and learning journal questions were adjusted accordingly. As a result, while the character of the questions remained consistent (i.e., ∼2-4 questions of various formats related to related content) they could cover a wider range of content. Sixteen videos (within fifteen separate learning journal assignments; ∼83% of the total learning journals) were used in the Fall 2015 semester and this number of videos and structure of learning journals was maintained for the Fall 2016 semester.

In-class activities – The incorporation of videos into pre-class assignments resulted in students being tasked with learning more content outside of class. Consequently, during class, several physical geology concepts could be addressed in more depth than during previous iterations of the course and/or more active learning activities could be incorporated into the class meetings using the time gained from moving some basic content delivery outside of the classroom. While many of these activities were piloted in the transitional semester of Fall 2015, the full suite of added in-class activities considered for this study were in place for the Fall 2016 semester. As a result, the Fall 2016 semester serves as the true treatment semester for the flipped class format as it involved the most complete combination of student work before class and in-class activities.

Measures

Student performance – Student performance was measured via three summative midterm exams. Each exam included either twenty-eight (Exams 1 and 2) or twenty-nine (Exam 3) dichotomously scored (correct or incorrect), multiple choice questions (85 in total) that were categorized at the lowest three levels of Bloom's taxonomy (knowledge, comprehension, application; Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Students also completed two short-answer questions that often involved concept sketches or open-ended questions that involved higher levels of Bloom's taxonomy (analysis, synthesis, evaluation) but these questions are not discussed here. Mean student performance for each question was calculated by summing the number of students who selected the correct answer and dividing this number by the total number of students who attempted the question for each semester. Kuder-Richardson Formula 20 (KR-20) values for each of the exams across the target semesters were 0.6 or greater, indicating relative reliability and internal consistency across students on each item of each exam (Kuder & Richardson, 1937). KR-20 is a special case of Chronbach's Alpha that describes the level of internal consistency for a measures that are dichotomously scored (i.e., right or wrong) and that contain items with a range of difficulty.

Student confidence – Beginning in the Fall 2014 semester, students indicated their level of confidence for each exam question response. Each question was followed by a horizontal line. The origin and terminus of the line were labeled "Not at all confident in my answer (0%)" and "Very confident in my answer (100%)," respectively (Figure 1). Students were instructed to place a mark along the line to represent their level of confidence that their answer choice was indeed the correct answer. The distance from the origin of the line to the student's intersecting mark was measured and converted to a percentage representing the student's confidence for each exam question (Figure 1). This process was repeated for each exam across the three semesters for which confidence measures were collected (Fall 2014 through 2016). Student confidence values were averaged for each question to generate a mean confidence elicited by that question across all students for each semester.

Exam question analysis –We analyzed the exam questions relative to how content was presented to students in Fall 2014 vs. Fall 2016 (e.g., reading vs. video, static in-class activities vs. augmented in-class activities, etc.). Each exam question was binned into one of five categories depending on how course procedures related to the topic changed across the study semesters (Table 2). Specific examples of each can be found in the Supplemental Materials.
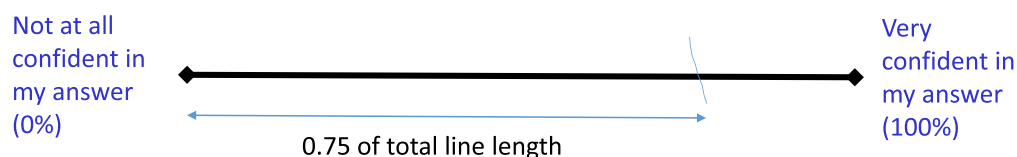
Figure 1. Example of data collection methods for student confidence used in the study. Lines were placed below each question of each exam.

Table 2. Exam question categories and explanations.

Category	n	Description
1	13	Question content removed from in-class presentation; first exposure to content moved to video
2	4	Question content augmented during in-class presentation
3	22	Question content first exposed to students in video AND in-class augmentation
4	14	Question content first exposed to students in video and static in-class coverage
5	32	No changes (i.e., static pre-class assignment and static in-class coverage

- **Category 1: Video only** - Exam questions in this category represented topics that were first introduced to students by video in a pre-class assignment and were subsequently excluded from the in-class presentation of content during the Fall 2016 semester. An example of this category would be questions related to the classification of igneous rocks, which can be found in the "Exam Question Analysis Examples" section of Supplemental Materials. These concepts were not further discussed in class beyond some diagnostic review questions at the start of the next lesson. This subsequently resulted in more class time that was devoted to discussion and explanation of more complex related concepts such as partial melting processes.

- **Category 2: Video + Same Lecture** - These exam questions assessed video-related content that was also covered in-class, but with the same level of detail as in previous semesters (i.e., constant in-class materials). The concepts were then reviewed in the classroom via the same presentation materials and level of detail as in Fall 2014 (i.e., no change in content delivered in-class).

- **Category 3: Video + New Activity** - This content was introduced to students in a required video and the instructor removed the related introductory content from the in-class lesson (similar to Category 1). The instructor then took advantage of the additional time made available during the resulting lessons to augment the in-class presentation of the content during the Fall 2016 semester. In short, the concepts these questions were designed to assess were introduced via a video in Fall 2016, but saw an augmentation of in-class instruction between Fall 2014 and Fall 2016.

- **Category 4: No Video + New Activity** - This content was not included in video resources but was part of a related reading assignment that was completed every semester. The content was augmented in lecture during Fall 2016 as a result of removing coverage of other foundational concepts to pre-class videos. There were relatively few ($n = 4$) examples of this exam question type but examples included questions related to hot spots, a topic that saw an added activity and the inclusion of multiple conceptest questions (conceptual multiple choice questions) as a result of newly-afforded time to cover the concept in greater depth in the classroom.

- **Category 5: No Change** - The final category represents exam questions that remained unchanged in both course coverage and content delivery throughout the four-semester period. These were questions that were related to content that was not communicated to students by a video and/or was not part of the pre-class assignments at any point in the study. Category question quantities and descriptions are provided in Table 2.

Theoretically, these five question categories represent the available changes allowed by partial flipping, with questions related to content removed from in-class presentation (Categories 1 and 2) leading to gained time for augmentation of that or other content (Categories 3 and 4). Together, looking across the target semesters of comparison (Fall 2014 vs. Fall 2016), the analysis of student performance and confidence was used to provide insight to the effects of classroom flipping on an introductory physical geology course at the college level.

Quantifying video design elements - To characterize and measure the design of online educational videos against the aforementioned principles of effective multimedia design, we developed a rubric to score Geoscience Videos on their design elements. This

required multiple phases of development. The rubric used for this analysis was generated after three iterative phases of design, revision and testing on publicly-available videos (not Geoscience Videos). The rubric consists of four items that were designed to assess the design elements of the rated video against the principles of multimedia design as suggested by the work of Mayer (2003). Each item is rated between 1-4 and video scores are summed to generate a total score (out of 16 points) that communicates the relative effectiveness of the video (with higher scores correlating to better-aligned videos). A full description of the rubric and the rubric itself is available in the Supplemental Materials for readers to use for their own decision-making in selecting videos to utilize in flipping activities.

To determine the ability to substantiate claims regarding effects of video implementation in the course over the target semesters, we had an individual external to the production of the videos (co-author Bedward) watch and rate fifteen of the sixteen videos on each item of the rubric to determine a total score for each video. Additionally, to investigate potential variability of ratings across raters, a second individual unrelated to this work also watched and rated a five of the videos used in the course on the rubric. Of these five videos, each rater returned a high average video score (14.6 & 13.8 out of a total 16 points). Additionally, the rater who rated fifteen of the videos used in the target course (Bedward), returned an average rating of 15.0 out of a possible 16. Full rubric description and ratings of each video by each rater is presented in the supplemental materials for this work.

Statistical methods

Quantitative data relating to student performance and confidence across the study semesters were analyzed using IBM Statistical Package for the Social Sciences (SPSS). All inferential statistics were run at an alpha level of .05. Effect size considerations for one-sample t-tests are reported as 95% confidence intervals, with larger interval distance from zero indicating an effect of larger magnitude. Effect size considerations for independent samples t-tests (d) follow recommendations from Cohen (1988), with sizes being defined as "small," ($d = 0.2$ - 0.49) "medium," ($d = 0.5$ - 0.79) and "large" ($d > 0.8$). Effect sizes for Mann-Whitney U and Wilcoxon signed rank tests (r) were calculated by dividing the test score (W) by the summed ranks of the sample size for each test (n; Fritz, Morris, & Richler, 2012; Kerby, 2014). Effect sizes for r were

considered "small" (0.1), "medium" (0.3), and "large" (0.5), also outlined in Cohen (1988).

Student performance

Initial statistical analysis of the question performance data across the four target semesters revealed that each distribution severely violated the assumption of normality (Shapiro-Wilk test for normality, p < .001 for each). Analysis of each distribution, however, revealed that each were similar in shape. Non-parametric Mann-Whitney U tests (two-tailed) were used to compare means of average student performance on questions across the three course exams for each study semester.

Student confidence

After calculating descriptive statistics for data pertaining to student confidence for each question included across the Fall 2014, 2015, and 2016 semesters, distributions failed to reject the null hypothesis for Shapiro-Wilk tests for normality and Levene's test for the assumption of equal variances so independent samples t-tests were performed to compare global mean confidence values across each semester.

Question category analysis

The paired data related to student performance and confidence for each question was divided into the five categories outlined above related to the changes made in the course between the target semesters. First, to determine whether there were any significant changes between the true (no video) baseline Fall 2013 semester and the subsequent Fall 2014 semester (8 videos pre-class but no changes to other course activities), mean performance on each question within each category in 2014 was subtracted from the mean performance on the same question in 2013 to generate a distribution of change scores for each category. Four of these five distributions failed to reject the null hypothesis of normality for the Shapiro-Wilk test (i.e., were normally distributed), so one-sample t-tests were performed on these data to determine if the mean was statistically significantly different from zero. One distribution (Category 1 Performance) was non-normal, so its dataset was subjected to a Wilcoxon signed ranks test to confirm or refute equivalency.

To determine the net change for each of the question-related variables across semesters during which the in-class activities were altered (Fall 2016 vs. Fall 2014), student performance and confidence for each

question in Fall 2016 was subtracted from the recorded values for the same question in Fall 2014 to generate a value representing the change in the variables between the two semesters. These delta values, grouped by question category, were used in statistical analyses. Six of the ten distributions [5 categories × 2 variables each (question performance, confidence)] failed to reject the null hypotheses for the Shapiro-Wilk test for normality (i.e., were normally distributed) and Levene's test for the assumption of equal variances with no significant outliers. The six distributions included Categories 1, 2, and 4 for question performance and Categories 2, 3 and 4 for confidence. Consequently, to determine if the change between semesters was significantly different, one-sample t-tests were performed on each distribution of comparison values. Additional information regarding statistical methods performed on each exam category distribution is available in supplemental materials.

Results

Comparing across the four semesters of the course analyzed in this study, there were several variables that remained consistent and others that improved in association with changes in content delivery and in-class activities across each iteration. Considering each set of question results in total across the four semesters, there were no statistically significant differences between overall student performance between the semesters as revealed by Mann-Whitney U tests, although mean performance in Fall 2016 was the highest of the study semesters (Table 3). Comparing student confidence across each semester as measured by mean confidence for each question indicated by students who answered that question, values were largely similar.

Mean overall exam performance was similar between Fall 2014 and Fall 2013 and there was also no statistical difference between student performance on the questions from each of the five categories between these semesters ($p > .05$; Table 4). This establishes equivalency between true control (2013) and first implementation of pre-class video and confidence-measuring (2014) semesters but with no significant alteration to in-class activities. Remaining results pertaining to category analysis between the Fall 2014 and Fall 2016 semesters are detailed in relation to each of the study's primary research questions.

How did the increase of video-related content delivered outside of class affect student performance and mean level of confidence on related content?

For questions assessing concepts that were no longer addressed during in-class activities or lectures (Category 1), there was no significant difference in student performance ($t(12) = 1.17$, $p = .26$), although the mean of the difference was positive (+2.1%; Figure 2; Table 4). This is in contrast to student confidence on these same questions, which showed a statistically significant decrease ($t(12) = -2.98$, $p = .01$, CI = -3.67, -0.57) of similar magnitude (-2.1%). For Category 2 questions relating to concepts introduced to students via a video in pre-class work with constant in-class activities between the two semesters, there was a significant increase in student performance ($t(13) = 2.30$, $p = .04$, CI = 0.36, 10.33), with students performing 5.2% better on these questions as compared to Fall 2014. Category 2 confidence showed a small increase (+1.29%) that was not statistically significant (Table 4).

How did the increased use of video-related content affect student performance and mean confidence related to concepts that were covered more extensively in class as a result of the flipped format?

Exam topics that were covered more extensively in-class as a result of moving foundational content to video represented questions in Category 3 and Category 4 (Figure 3). Questions that were related to

Table 3. Mean and standard deviations for question performance and confidence ($n = 85$).

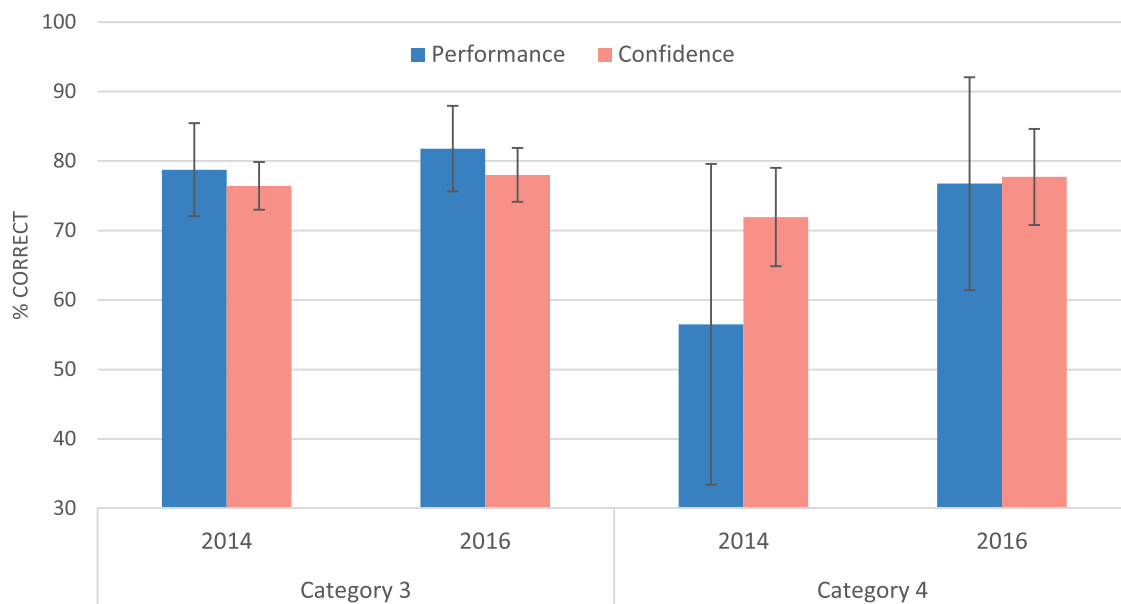|	Fall 2013		Fall 2014		Fall 2015		Fall 2016	
	M	SD	M	SD	M	SD	M	SD
Perf.	77.48	16.51	77.06	16.02	77.51	18.05	80.97	15.95
Conf.	-	-	76.86	7.41	75.25	8.20	77.86	8.54

Table 4. Means and standard deviations of category change scores.

			Category 1	Category 2	Category 3	Category 4	Category 5
F2014-F2013	Performance	M	−1.82	-3.92	2.41	−2.47	−2.01
		SD	7.18	10.66	7.89	9.47	7.29
F2016-F2014	Performance	M	2.12	20.25	2.82	5.20	3.36
		SD	6.52	10.64	9.64	8.44	7.43
	Confidence	M	−2.12	5.77	1.58	1.29	0.78
		SD	2.56	1.99	4.29	4.93	4.09
							*$p < .05$
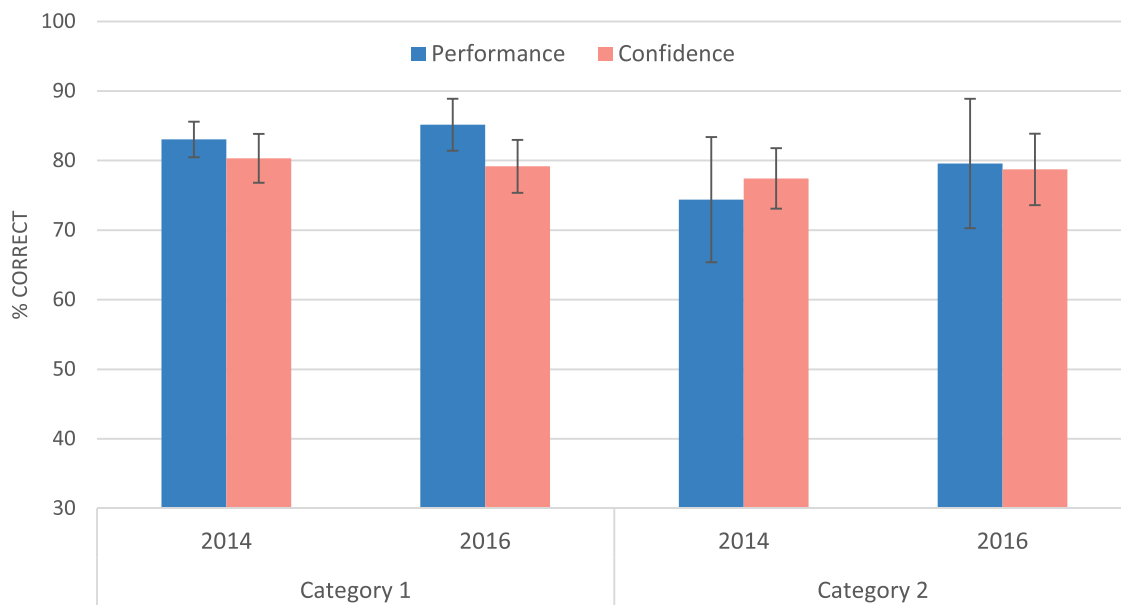
Figure 2. Mean percentage of correct responses for questions from Category 1 and Category 2 for the Fall 2014 and Fall 2016 semesters. There is no statistical difference in performance in questions from Category 1, though confidence was significantly less (−2.1%). Performance on Category 2 questions were determined to be significantly higher in 2016 ($t(13) = 2.30$, $p < .05$). Error bars represent 1.96 SE.

Figure 3. Mean percentage of correct responses for questions from Category 3 and Category 4 for the Fall 2014 and Fall 2016 semesters. Performance and confidence significantly increased in Category 4 questions and did not change in Category 3 questions. Error bars represent 1.96 SE.

concepts first introduced via video and subsequently augmented during in-class meetings were not significantly different in Fall 2016 than Fall 2014 ($t(21) = 1.37$, $p = .18$, CI = -1.45, 7.09), although students did perform better on average in 2016 (M = 2.82 SD = 10.10). Confidence for Category 3 questions also increased (M = 1.58 SD = 4.29), but not at a statistically significant level ($t(21) = 1.73$, $p = .10$, CI =

-0.32, 3.48; Figure 3, Table 4). Category 4 questions featured concepts that were augmented in-class but were not introduced to students via a video assignment. There was a significant increase in student performance ($t(3) = 3.81$, $p = .03$, CI = 3.32, 37.18) despite a small sample size ($n = 4$) with students scoring on average 20% better on these questions. Measured student confidence also significantly

Figure 4. Mean percent change for each question category between treatment and control semesters (Fall 2016—Fall 2014). Symbol (*) indicates significant change from zero at $p < .05$ level.

increased for these questions as mean confidence increased by nearly 6% (M = 5.76 SD = 1.99; $t(3)$ = 5.77, $p = .01$, CI = 2.60, 8.94).

How did students respond to questions on which there was no change in presentation?

Finally, in response to questions in Category 5, which had no changes in either pre-class content delivery or in-class activities, students performed significantly better in Fall 2016 than Fall 2014 ($t(31) = 2.56$, $p = .02$, CI = 0.68, 6.04), increasing their performance on these 32 questions by an average of 3.36% (SD = 7.43). Confidence for questions in Category 5 was unchanged between the semesters with a Wilcoxon signed rank test revealing a difference insignificant from zero ($W = 359$, $p = .08$, $r = .68$). A summary of all category change scores and associated statistical significance is provided (Figure 4).

Discussion

Major findings

This study sought to isolate the effects of utilizing short, content-related videos to move foundational concepts in physical geology outside of the classroom, thus freeing up time during in-class meetings for deeper content coverage and active learning techniques. Results suggest a nuanced relationship between variables related to classroom flipping and one that, in many ways, highlight that the true "effects" of partial course flipping must be delineated by the goals of the instructor implementing the course changes. Is the

goal to move presentation of material outside of the classroom? Or is the goal to increase student learning? The strong relationship between active learning strategies and student performance in higher education environments suggests that these pursuits must be intertwined for course flipping to truly benefit student outcomes in undergraduate STEM-related courses.

Though not statistically significant, students overall performed better in the Fall 2016 semester than in prior semesters of the course. There are statistically significant changes in student performance and confidence when we break down the summative exam questions by category related to how content was presented both prior to class and during class. For concepts introduced in videos and then removed from in-class presentation (Category 1), students maintained the same level of performance despite specific concepts no longer being presented by the instructor during class. Consequently, we infer that students are capable of teaching themselves some of the basic course content, provided that the multimedia they are consuming is well-aligned to empirical design suggestions (as outlined in Mayer, 2003). Though performance was maintained across the semesters for questions pertaining to these concepts that were removed from in-class presentation, there was a slight drop in student confidence. This suggests a need for more cognitive and metacognitive support for these topics throughout the course, perhaps through more formative assessment and feedback regarding how students are performing on concepts that were no longer covered in the course lectures.

Whereas performance on topics introduced in videos and removed from lecture remained constant, performance on questions related to concepts introduced in videos and subsequently reinforced via in-class presentations and activities (Category 2) saw a significant increase between the comparison semesters. Perhaps videos have greater potential than readings to effectively engage students as they gather prior knowledge and generate their own representations of the concepts that they subsequently bring to class to serve as a foundation for instruction. The added value of supporting learning with videos is reinforced by the work of Mayer (2003), and with recent findings on the effectiveness of geoscience videos specifically in fostering learning of geology content. Wiggen & McConnell (2017) determined in a laboratory study that videos on the physical geology concepts of magma viscosity and fault identification were more effective in conveying information than equivalent reading passages from geoscience textbooks as measured by pre- and post-tests on the content. One such question included in Exam 1 of this study but related to the content covered in the video investigated in Wiggen & McConnell (2017) saw over a 23% improvement in performance (and 10% more confidence) when comparing performance on the question between Fall 2016 and Fall 2014. These results support video usage in this context.

There were mixed results for exam questions related to topics that were augmented in lecture. One category of questions, whose content was not introduced to students in videos (Category 4), exhibited large improvements in both student performance and confidence. Additionally, for these questions (as they are few), it is possible to isolate specific strategies that were employed during in-class meetings that likely helped contribute to the exhibited gains. Two of these questions were related to the concept of hot spots, with in-class augmentations on the topic in Fall 2016 coming in the form of additional conceptual multiple choice questions that were afforded via the flipping process. This provides support to prior studies that claimed benefits seen in flipped courses may merely be those of active learning strategies (Jensen et al., 2015). Importantly, however, this effect was not homogenous across questions with augmented in-class activities. The results of exam questions assessing video topics that saw more depth or an increase in activities in class for the Fall 2016 semester (Category 3) were not significantly different than in Fall 2014 (although results increased on average). This suggests that even though pre-class videos can be more effective than text-based assignments in communicating requisite knowledge (as evidenced by results for questions in Categories 1 and 2), the diverse nature of the discipline, of the strategies that can be used in-class to communicate the discipline, and of students in general dictates that there may be a differential effect of these variables on the learning of geology content. The video used to communicate baseline knowledge, the activities used to build upon this foundation, or receptivity of the student may all affect student learning. Others have reported that not all multimedia will have an equal benefit to learning as measured by an assessment (Chen et al., 2010). Further investigation into this line of inquiry and further isolation of effective strategies relating to specific topics of the discipline (e.g., fault mechanics, lithospheric thickness at different tectonic settings, etc.), however, is an important avenue for future work.

Finally, results from questions in the final category of analysis (Category 5) suggests that in spite of course design elements, there will always be exam questions and course concepts to which students will perform differentially. Although these questions saw a small yet significant mean increase comparing Fall 2016 and Fall 2014 (3%), it is difficult to speculate to gains measured in these questions as no variables were altered in reference to these concepts within this study. Perhaps students in the two courses simply started with different levels of prior knowledge. However, it may be that students were able to devote more time and attention to studying and learning these concepts because of increased efficacy for the remainder of course content.

Considering the effects of video-driven flipping on student confidence across the target semesters, there was relatively little change as a result of the course alterations, even when significant changes in performance were present. Only two of the five question categories saw a statistically significant change in student confidence, and both of these fluctuations were no greater than approximately 5%. This is in contrast to prior work in a laboratory setting investigating the use of geoscience videos and their efficacy on this variable. Wiggen & McConnell (2017) demonstrated more-ubiquitous and significant increases in student confidence related to questions addressing content learned either by watching a geoscience video or after reading an equivalent passage of text (Wiggen & McConnell, 2017). While participants who watched a video demonstrated higher confidence than participants who read the textbook passage, all demonstrated significantly increased confidence in their answers.

This suggests that student confidence may be potentially more difficult to impact in a situated (i.e., real-world course) setting than performance and future work related to the malleability of confidence judgments in these settings is suggested.

This study was designed to help further identify the effects of classroom flipping on student learning in introductory science courses. As the research designs for such studies have seen wide variability, we attempted to constrain many of the situational factors facing the two semesters being studied to further identify how college students respond to the course alterations over the course of a semester. When comparing these results to work from other studies investigating flipping in undergraduate STEM settings, similar to Jensen et al. (2015) we found similarity in student performance on topics whose background material was removed from in-class consideration (e.g., Category 1). We also, however, saw increases in student performance in questions addressing topics that were augmented with active learning via newly afforded class time (Category 4; similar to Gross et al., 2015). Additionally, the increase in performance on questions that saw no in-class change Category 2, however, suggest potential benefits from the adoption of video vs. text (as suggested by Wiggen & McConnell, 2017). While promising but definitely not clear-cut, these results contribute evidence to the literature base that the practice of removing introductory concepts from in-class instruction, if accompanied with targeted videos and homework assignments, does not equate to removing the concepts from the course altogether (i.e., students can learn them outside of class). Additionally, the practice of flipping can afford new time for in-class activities related to more-difficult concepts to potentially increase the opportunity for student success.

Limitations

Although this study did find support for partial classroom flipping at the college level, there are several limitations to consider in the interpretation of results. Though multi-semester-long studies conducted in situated learning environments can provide important insight into student learning in a real-world sense (as opposed to laboratory settings), increasing external validity, there are several variables that cannot be controlled. Although this study attempted to control many of the situational factors related with the course setting (instructor, time of year, time of day, exams, etc.) implications and notions of causation should be approached with caution. Additionally, although each dataset was relatively robust

in terms of its origin (∼75-95 student responses for 85 questions across four semesters), the categorization of questions cannibalized the statistical power of the tests. Though many results of the study's statistical tests were still significant, it is important to consider these smaller samples in the interpretation of results.

As students were not given pre-measures or post-measures at the beginning and end of each semester, gains seen during the Fall 2016 semester could simply be due to the sample population of students having increased prior knowledge of the discipline or increased motivation for learning science and/or geology. Also, for future work, other variables such as incoming GPA and course attendance could further isolate effects. Unfortunately, these data were not available for this study. Additionally, the target course, even in its Fall 2013 iteration, is one that has been designed to include a suite of research-based teaching practices and student-centered pedagogy and as such there may be a ceiling effect for student performance in the course (mean exam performance for Fall 2016 was above 80%). Though students performing "too well" is certainly not a limitation in itself, it is important to consider this relatively high performance in the consideration of results. Future work will seek to control for prior knowledge and teaching practices to better isolate effects of pedagogical interventions on the target variables.

Finally, though many of the features of the course were equivalent across the four iterations of the course, students were not subject to identical homework assignments or formative assessment support, introducing the potential for confounding variables relating to student performance and confidence in the course. Platforms offering students optional formative quizzes was experimentally altered for a concurrent project investigating student metacognitive monitoring accuracy (Jones & McConnell, 2016), which also may have potentially influenced student performance on exam questions during the Fall 2016 semester. However, all other features related to the role of the online quizzes in the course were left constant (i.e., they were optional for both semesters and not considered for a grade). Given the modest net differences between the Category variables and the similarity of the other situational factors regarding the course (e.g., Fall semester, same time, same population of students, etc.), this influence was likely minimal.

Conclusion and suggestions for instructor practice

The future of undergraduate education is likely to feature a greater reliance on teaching with online

resources regardless of whether students will be enrolled in face-to-face or hybrid or fully online courses. This study demonstrates that student learning of basic content using online resources that incorporate short videos can be just as effective as an instructor-guided face-to-face lesson. The target course for this study utilized videos that adhered to empirically-supported aspects of multimedia design and used time gained from the flipping process to utilize active learning strategies during in-class meetings. It is our experience, however, that effective flipping practices should add depth, complexity, and student activity to the target course, not merely shift static actions to different settings. It is suggested that future investigation into the effect of flipping on student performance should include further controls of student-level variables such as prior knowledge, motivation, video design, and multimedia consumption, and further strive to identify the best balance of in-class active learning strategies and out-of-class interactive multimedia learning.

Instructors can move some basic content learning outside of their class meetings to gain time for in-class activities. While planning this transition, however, we suggest that instructors select material that is not going to be too challenging for students to comprehend. For example, generating pre-class homework assignments that ask students to classify rocks, faults, volcanoes or other geologic features. We created a video on why and how we flipped our class (see, https://youtu.be/1tBhm8uBkhM) that may be useful for some readers.

There are many videos available online and students tell us that they routinely seek out what they think are useful videos to support their learning (Wiggen and McConnell, 2017). Instructors may assess potential videos to incorporate into their courses using the rubric (see Supplemental Materials) to ensure effective multimedia design. The Geoscience Videos YouTube channel (www.youtube.com/c/geosciencevideos) contains 35 short geoscience videos (see list in Supplemental Materials), many of which we have used across flipped face-to-face and blended versions of our introductory physical geology course.

We also counsel care in considering the length of a video or other "flipping" resource. A single 6-7 minute video can provide material that would take you approximately 15-20 minutes of in-class instruction. This would typically be combined with several related questions offered through the course CMS. For our purposes, this struck a reasonable balance between student responsibility for their learning and the amount of additional time that we could free up to incorporate additional in-class activities. McConnell et al. (2017) describes a variety of active learning strategies including examples from the geosciences. We created a blog to parallel the Geoscience YouTube channel that summarizes how we adapted our in-class teaching as we adapted the flipped class format (see https://geosciencevideos.wordpress.com/).

Do not expect ubiquitous success in the process. While this study saw some gains in topics that were either communicated through video or augmented during additional in-class activities, there were others questions that did not see relative gains despite targeted interventions. The instructors applying the flipped model should be reflective practitioners and not be afraid to experiment with the details of the process to suit the needs of their students and their teaching environment.

Acknowledgments

Funding

ORCID

Jason P. Jones http://orcid.org/0000-0001-6563-8185
David A. McConnell http://orcid.org/0000-0003-4879-9520
Jennifer L. Wiggen http://orcid.org/0000-0002-2360-1963
John Bedward http://orcid.org/0000-0003-3202-8713

References

Baddeley, A. D. (1999). *Human memory*. Boston, MA: Allyn & Bacon.
Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
Bennett, P., & Glover, P. (2008). Video streaming: Implementation and evaluation in an undergraduate nursing program. *Nurse Education Today*, 28(2), 253–258. doi:10.1016/j.nedt.2007.04.005

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

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals: Handbook I: cognitive domain* (No. 373.19 C734t). New York, NY: D. Mckay.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning, creating excitement in the classroom* (ASHE-ERIC Higher Education Report No. 1). Washington, DC: The George Washington University, School of Education and Human Development.

Burchfield, C. M., & Sappington, J. (2000). Compliance with required reading assignments. *Teaching of Psychology, 27*, 58–60.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293–332. doi:10.1207/s1532690xci0804_2

Chen, Z., Stelzer, T., & Gladding, G. (2010). Using multimedia modules to better prepare students for introductory physics lecture. *Physical Review Special Topics, Physics Education Research, 6*, 1–5.

Choi, H., & Johnson, S. (2005). The effect of context-based video instruction on learning and motivation in online courses. *The American Journal of Distance Education, 79*, 215–227. doi:10.1207/s15389286ajde1904_3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: ten years of experience and results. *American Journal of Physics, 69*(9), 970–977. doi:10.1119/1.1374249

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*(1), 4–14. doi:10.1016/j.learninstruc.2012.06.001

Dixon, J. L. (2016). Communicating geoscience concepts in the classroom and beyond using short video-based resources. *Geological Society of America Abstracts with Programs, 48*, 7. doi: doi:10.1130/abs/2016AM-279486

Dixon, J. L., & McConnell, D. A. (2014). Geoscience videos and their role in supporting student learning in hybrid introductory geoscience courses. *Geological Society of America Abstracts with Programs, 46*(6), 735.

Dixon, J. L., McConnell, D. A., & Bedward, J. (2015). Communicating geoscience concepts beyond the classroom: Who is watching geoscience videos on YouTube? *Geological Society of America Abstracts with Programs, 47*(7), 837.

Evans, D. (2011). Turning lessons upside down. *Times Educational Supplement, 4*.

Fink, L. D. (2003). *Creating significant learning experiences*. San Francisco, CA: Jossey-Bass.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. doi:10.1073/pnas.1319030111

Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education, 10*(2), 175–186. doi:10.1187/cbe.10-08-0105

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2–18. doi:10.1037/a0024338

Gajjar, N. B. (2013). The role of technology in 21st century education. *Int. J Res. Educ, 2*, 23–25.

Green, S. M., Voegeli, D., Harrison, M., Phillips, J., Knowles, J., Weaver, M., & Shephard, K. (2003). Evaluating the use of streaming video to support student learning in a first-year life sciences course for student nurses. *Nurse Education Today, 23*(4), 255–261. doi:10.1016/S0260-6917(03)00014-5

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *Cell Biology Education, 14*(4), 1–8.

Heyborne, W. H., & Perrett, J. J. (2016). To flip or not to flip? Analysis of a flipped classroom pedagogy in a general biology course. *Journal of College Science Teaching, 45*(4), 31–37.

Jensen, J. L., Kummer, T. A., & Godoy, P. D. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE Life Sciences Education, 14*, 1–12.

Jha, V., Widdowson, S., & Duffy, S. (2002). Development and evaluation of an interactive computer-assisted learning program - a novel approach to teaching gynecological surgery. *British Journal of Educational Technology, 33*(3), 323–331. doi:10.1111/1467-8535.00267

Jones, J. P., & McConnell, D. A. (2016). CLASS: A new tool for characterizing student awareness of their learning in geoscience courses. *Geological Society of America Abstracts with Programs, 48*, 7. doi:10.1130/abs/2016AM-279609

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. doi:10.1207/S15326985EP3801_4

Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology, 3*, 11.IT.3.1. doi:10.2466/11.IT.3.1 doi:10.2466/11.IT.3.1

Kortz, K. M., Grenga, A. M., & Smay, J. J. (2017). Establishing and applying literature-based criteria for effective communication of science to novices via introductory geology textbooks. *Journal of Geoscience Education, 65*(1), 48–59. doi:10.5408/16-205.1

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika, 2*(3), 151–160. doi:10.1007/BF02288391

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *Journal of Economic Education, 31*(1), 30–43. doi:10.2307/1183338

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41*, 85–139.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*(2), 125–139. doi:10.1016/S0959-4752(02)00016-6

Mazur, E. (1997). *Peer instruction: A user's manual.* Upper Saddle River, NJ: Prentice Hall.

McConnell, D. A., Chapman, L., Czajka, C. D., Dixon, J., Jones, J., Ryker, K. D., & Wiggen, J. (2017). Instructional utility and learning efficacy of common active learning strategies. *Journal of Geoscience Education*, 65(4), 604–625. doi:10.5408/17-249.1

McCrindle, A. R., & Christensen, C. A. (1995). The impact of learning journals on metacognitive and cognitive processes and learning performance. *Learning and Instruction*, 5(2), 167–185. doi:10.1016/0959-4752(95)00010-Z

Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before lecture: A strategy that improves learning outcomes in a large introductory biology class. *Cbe—Life Sciences Education*, 9(4), 473–481. Winter). doi:10.1187/cbe.10-04-0063

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*. 8, 1–28.

Podolefsky, N., & Finkelstein, N. (2006). The perceived value of college physics textbooks: students and instructors may not see eye to eye. *The Physics Teacher*, 44(6), 338–342. doi:10.1119/1.2336132

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. doi:10.1007/s11409-008-9031-3

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36(1–2), 111–139. doi:10.1007/s11165-005-3917-8

Schultz, D., Duffield, S., Rasmussen, S. C., & Wageman, J. (2014). Effects of the flipped classroom model on student performance for advanced placement high school chemistry students. *Journal of Chemical Education*, 91(9), 1334–1339. doi:10.1021/ed400868x

Seeling, P., & Reisslein, M. (2005). Video in distance education: ITFS vs. web-streaming: Evaluation of student attitudes. *Internet and Higher Education*, 8, 25–44. doi:10.1016/j.iheduc.2004.12.002

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (Eds.). (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.

Stelzer, T., Gladding, G., Mestre, J. P., & Brookes, D. T. (2009). Comparing the efficacy of multimedia module with traditional textbooks for learning introductory physics content. *American Journal of Physics*, 77(2), 184–190. doi:10.1119/1.3028204

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, 15(2), 171–193. doi:10.1007/s10984-012-9108-4

Tantrarungoroj, P. (2008). *Effect of embedded streaming video strategy in an online environment on the learning of neuroscience* (Doctoral dissertation). Indiana State University, Terre Haute.

Tucker, B. (2012). The flipped classroom. *Educ. Next 12*. Retrieved from http://educationnext.org/the-flipped-classroom

Tune, J. D., Sturek, M., & Basile, D. P. (2013). Flipped classroom model improves graduate student performance in cardiovascular, respiratory, and renal physiology. *Advances in Physiology Education*, 37(4), 316–320. doi:10.1152/advan.00091.2013

Wetzel, C., Radtke, P., & Stern, H. (1994). *Instructional effectiveness of video media*. Hillsdale, NJ: Lawrence Erlbaum.

Wiggen, J., & McConnell, D. (2017). Geoscience Videos and Their Role in Supporting Student Learning. *Journal of College Science Teaching*, 46(6), 44–49.

Zhang, D., Zhou, L., Briggs, R., & Nunamaker, J. Jr., (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management*, 43(1), 15–27. doi:10.1016/j.im.2005.01.004

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. doi:10.1207/s15326985ep2501_2

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166–183. doi:10.3102/0002831207312909