CAN VIDEO EXPERIMENTS BE AS EFFECTIVE AS HANDS-ON EXPERIMENTS FOR CONCEPTUAL LEARNING?

Margot Vigeant and Amy Golightly
Bucknell University • Lewisburg, PA 17837

INTRODUCTION

By necessity in late Spring 2020 (Sp20), many courses designed for in-person instruction at US institutions suddenly became remote courses. The mode of emergency remote instruction for laboratory courses posed a particular challenge, and a variety of approaches were attempted with the acknowledgement that instructors were all doing their best to replace an irreplaceable experience. It appears that a blend of in-person and remote learning for laboratories will continue at some level for many campuses, at least for the near future. Since instructors now have a little more time with which to plan these experiences, this study aims to aid instructional decision-making as faculty choose the modes of alternative instruction in which to invest their time. The question driving this research is: To what extent does the video of an experiment replicate the educational outcomes of the hands-on experiment in inquiry-based learning of chemical engineering concepts? While the case described below is for chemical engineering thermodynamics, the results are suggestive of the utility of video replacement for any short laboratory experience centered on teaching physical concepts.

BACKGROUND

Student misconceptions about science and engineering concepts can be surprisingly resilient to lecture-based instruction.[1,2] While lecture, examples, and practice-and-feedback in the form of homework and other assessments are effective for sharing declarative and procedural scientific and engineering knowledge — such as the definitions of terms and the applications of equations — proper understanding of concepts is less likely to result from these approaches. One approach that can be used effectively for conceptual learning is inquiry-based laboratory activities (IBLAs).[3] These activities use a surprising or “discrepant” event in which a misconception-based prediction is not borne out by real-life events to create a teachable moment in which students can repair their conceptual understanding.[3]

Such discrepant events do not automatically result in learning; in fact, work has shown that students are capable of ignoring results that fail to conform to their prior understanding.[4,5] One approach to make IBLAs effective is to use writing, reflection, peer instruction, and guided inquiry to make the most of the situation.[6] For example, students should first commit in writing to their prediction about what is going to occur, then they should conduct the experiment, and finally they should interact with peers and faculty and craft a written reflection on how their understanding has been changed. This approach is illustrated in Figure 1 and was used in this work.
Earlier application of the IBLA approach in chemical engineering heat transfer and thermodynamics relied upon hands-on experiments for the discrepant event whenever possible, and simulations when practical experimentation was impossible.\textsuperscript{[7–9]} When these IBLAs were shared with other instructors, the research team discovered that even short hands-on activities can be difficult to implement in some settings, particularly in larger courses without laboratory sections. A subset of the IBLAs within heat transfer was selected for rewriting in one of four additional presentation modes: demonstration, simulation, demonstration-of-simulation, or thought experiment. Each approach maintained the same pre-activity prediction questions and post-activity reflection questions. However, the “interesting discrepant event” was delivered in one of the five aforementioned modes. It was hypothesized that the move away from direct, hands-on experience might render the activities less convincing (and therefore less effective) agents of conceptual change. On the other hand, it was thought that perhaps maintenance of the prediction questions and reflection questions might preserve the creation of the “teachable moment” and consolidation of understanding that comes from reflection. In the end, the work demonstrated that while hands-on activities generated the largest gains in conceptual understanding, all other modes were significantly better than teaching without IBLAs.\textsuperscript{[10–12]} A thought-experiment-like approach of analogies also provided some conceptual gains,\textsuperscript{[13]} although also lower than other modes.

Unfortunately, the instructional modes described above require a level of preparation that might not be possible during emergency remote instruction. Programming a simulation, particularly one with graphics that accurately represent the experiment, requires time. Providing students with the experimental data required for the thought-experiment approach requires access to the laboratory. One mode of IBLA delivery that had not been previously studied was the use of video. In this mode of instruction, students complete pre-labs as normal and then in place of the hands-on experiment, they watch a video of the instructor completing that experiment. Afterwards, students write up their results and reflection as they otherwise would have done.

Experiments in physics instruction have suggested that live, in-class demonstrations are only slightly effective learning tools; something that may be improved with accompanying written reflections.\textsuperscript{[14]} Intriguingly, it appears that recorded video demonstrations may be more effective than live demonstrations,\textsuperscript{[15]} perhaps because the video allows instructors to highlight key elements and students to review what happened. Taken with the previous findings on IBLAs, this suggests that replacing hands-on activities with videos of those activities is likely to be at least as effective as demonstration.

Our question is: when the hands-on experiment is replaced with a video of that same experiment, is the educational outcome equivalent between the two modes, or is one better?

\section*{MATERIALS AND METHODS}

\subsection*{Concept Inventory}

The Concept Inventory for Engineering Thermodynamics (CIET) was used to assess conceptual learning and change in the course.\textsuperscript{[16]} This instrument has 35 questions and is available through the AIChE Concept Warehouse and contains both original questions and questions from the Thermal and Transport Concept Inventory.\textsuperscript{[17]} It addresses five concept areas as shown in Table 1. Earlier work with this concept inventory established an overall reliability of KR20 = 0.80 (n = 199), which is good for use as a research instrument. The reliability for the five-question RXN subscale was KR20 = 0.59, meaning its results should be interpreted carefully. Further work suggests that the lower reliability of this sub-section might be linked to the fact that a portion of the students in the test population took thermodynamics courses in which reaction thermodynamics was unaddressed.

\subsection*{Instructional Approach}

A sample of convenience consisting of two student cohorts, Sp19 and Sp20, in the Chemical Engineering Thermodynamics course was used. The Spring 2019 (Sp19) course had a
Before the switch to remote instruction, students were taught the Concept CIET in both the first and last weekly laboratory sessions of the semester; for the purposes of this study, identifiers were removed and only those who gave consent on both the pre- and post- administrations of the concept inventory were included in the analysis.

Both versions of the course used the Moodle learning management system (LMS). In this interface the materials for the course (reading, homework, test objectives, etc.) were arranged in blocks containing all of the materials for a given week. In all of Sp19 and the first part of Sp20, students were using the LMS regularly, and for laboratories in particular, the LMS housed the turn-in link for the post-lab write up as well as the link to the simulation for those IBLAs that used one. After week seven, the Sp20 cohort moved to emergency remote instruction. In order to accommodate students across time zones and living situations, the course became largely asynchronous, with a video-lecture used to set up the day’s example problem, and a quiz offered through the LMS to collect students’ worked answers to that problem, supported with faculty office hours. Five of the nine IBLAs were complete by the switch to emergency remote instruction. Two of the remaining IBLAs were for the “equilibrium vs. steady state” (Eq vs SS) concept area. The other two remaining IBLAs were for the “reaction rate vs. reaction equilibrium” (RXN) concept area. The first of the “Eq vs SS” IBLAs is a simulation and was shared with the students unchanged from its presentation in Sp19. The remaining three IBLAs were presented as hands-on experiments in Sp19 and did not have equivalent simulations available for remote instruction. In fact, the tactile nature of these particular experiments had discouraged their reproduction as simulations in the past. These three experiments were conducted by the instructor in her home kitchen, filmed and edited with Clips (Apple) on an iPhone®, and subsequently posted on YouTube™ and linked to the class LMS. Presentation of laboratories and, indeed, all class activities through the LMS, became more detailed during remote instruction. While the materials were still arranged in week-long blocks, additional class materials such as lecture videos and online reading/watching-check quizzes were embedded in a daily to-do-list format. That is, there would be a heading that said “Wednesday” and then under that, all of Wednesday’s class and lab materials in the expected order of completion. The lab handout was presented this way in the LMS with a note that students should complete the prediction questions before continuing, and then a link to the video (which also notes students should complete predictions before continuing). Finally, there was a link to turn in the write up. In the LMS, students were instructed to use the laboratory handout as though they were conducting the experiment and to make predictions before watching the video, watch the

<table>
<thead>
<tr>
<th>Concept Area</th>
<th>Abbreviation</th>
<th>Explanation</th>
<th># Questions in CIET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reversibility</td>
<td>REV</td>
<td>Students often assume reversible systems are easy/practical to create/use.</td>
<td>6</td>
</tr>
<tr>
<td>Entropy and the 2nd Law</td>
<td>ENT</td>
<td>Students underestimate the impact of entropy on engines/cycles, believing it’s generally possible to obtain very close to 100% efficiency.</td>
<td>9</td>
</tr>
<tr>
<td>Equilibrium vs. Steady State</td>
<td>Eq vs SS</td>
<td>Students believe that the two terms describe identical system states.</td>
<td>9</td>
</tr>
<tr>
<td>Internal Energy vs. Enthalpy</td>
<td>U vs H</td>
<td>Students confuse internal energy and enthalpy.</td>
<td>6</td>
</tr>
<tr>
<td>Reaction Rate vs. Extent of Reaction</td>
<td>RXN</td>
<td>Students expect that any highly favorable reaction (large negative delta-G) will automatically be a very rapid or even explosive reaction.</td>
<td>5</td>
</tr>
</tbody>
</table>
video, and then complete the reflection questions as normal. The student handouts for all three of these IBLAs are available at: http://www.projects.bucknell.edu/LearnThermo/index.html.

Videos

For each of the three IBLA videos, the goal was to capture some of the visceral reaction and surprise that students might otherwise have experienced in-person. For the “Eq vs. SS” IBLA “Hot Pot,” the surprise is that the metal handle of a pot of boiling water is not itself at 100 ºC (thermal equilibrium with the water temperature), but has a lower steady-state temperature that makes it safe to touch. The memorable surprise in the hands-on experiment is students being invited to lift the pot by the handle, should they feel comfortable doing so. In the video the instructor makes a number of temperature measurements inside and outside of the pot and on the handle before safely demonstrating that it is comfortable to hold. This is the longest video, at 1 min 47 sec.

The first IBLA for the RXN concept area is “Explosive Reactions?” where students determine the significant favorability of reactions in which iron, carbon, or silicon combine with oxygen. In the hands-on experiment the “surprise” comes in two parts — first when a production is made of “exposing carbon to oxygen” for students just to see that they’ve been given a pencil to wave around in the air. However, students are also generally surprised when extra fine steel wool, which contains both iron and carbon, is flammable. Both the non-event of iron and carbon exposed to oxygen and the red-hot glow of the ignited steel wool are captured in the video. This video is slightly longer than one minute.

The final IBLA is “Volcano” and is based on the reaction between baking soda (sodium bicarbonate) and vinegar (acetic acid). The surprise in this activity is primed by the calculation of the standard state Gibbs free energy change for the reaction, which makes the reaction appear unfavorable. The reaction does, nevertheless, occur at room temperature. In class it is common for many of the students to have prior experience with this reaction, so the honor of adding the baking soda to the vinegar is usually given to a student who did not perform this reaction as a child. The video is 49 seconds long and shows the straightforward mixing of the two components in a small glass jar, where the resulting CO2 bubbles generated by the reaction may be clearly observed. Figure 2 shows still images from the video portion of the three IBLAs that were altered for emergency remote instruction in Sp20.

Analytical

Sixteen respondents from Spring 2019 and 19 from Spring 2020 who gave consent and completed both pre- and post-tests were included in the analysis. The 2019 sample was comprised of 9 males and 7 females, and the 2020 sample was comprised of 10 males and 9 females. The 2019 cohort had 13 who identified as white and 3 who identified as Asian/Pacific Islander, and the 2020 cohort had 15 who identified as white, 1 who identified as Hispanic/Latino, and 3 who identified as Asian/Pacific Islander. Differences between the cohorts in terms of gender or race/ethnicity were not statistically significant.

On the CIET, student responses were dichotomized into correct and incorrect responses, and scores were considered both overall (35 questions) and on the REV, ENT, U vs. H, and RXN area subscales. The RXN subscale was selected for separate analysis because it allows direct comparison of a concept addressed exclusively by experiment in one cohort and exclusively with video in the other. The “Eq vs SS” subscale was not considered separately because the concept was addressed by a mix of approaches (simulation, experiment, video) in both cohorts. SPSS v26 was used to analyze the data. Pre-test total scores for both cohorts were compared with an independent samples t-test, and no significant differences were found with 2019 (n = 16) mean = 20.06; σ = 4.34 and 2020 (n = 19) mean = 19.00; σ = 5.86; p > .05. These findings suggest that students’ understandings of concepts were relatively consistent across cohorts. To examine gains made by each cohort, paired samples t-tests were conducted.
To examine the pragmatic significance of the findings, Cohen’s $d$ was calculated for overall and reaction subscales for both cohorts.

**RESULTS AND DISCUSSION**

Concept inventory results, overall and for the subscales are summarized in Table 2. Three subscales capture concepts where activities were the same for both cohorts (ENT, REV, and U vs. H), one subscale with one activity the same and the other either hands-on (Sp19) or video (Sp 20, “EQ vs SS”), and one sub-scale (RXN) where all activities were either experiment (Sp19) or video (Sp20). Students demonstrated both statistical and pragmatically significant improvements on overall score between the pre- and post-test in both cohorts. However, the improvement and the corresponding effect size are greater for students in the hands-on IBLA group.

Note that the pre-test scores do not show significant differences between the two cohorts initially, which was expected based on their similar prior coursework. There was no significant change in understanding for either cohort for ENT and REV concepts from the beginning to the end of the semester, in contrast to prior cohorts where improvement was seen. In the “EQ vs SS” subscale, it is not possible to disambiguate the questions most influenced by the two different instructional approaches experienced in the course of that topic; these results are shared for completeness and cannot be interpreted to support or refute the effectiveness of video-based experiments.

These findings suggest that while both groups made statistically and pragmatically significant gains from the beginning to the end of the semester in terms of their overall concept inventory score, the students who participated in in-person instruction showed gains almost a half a standard deviation (0.46) higher than their peers, who switched to online education midway through the semester. Viewing in terms of percent correct, the 2019 cohort started with a pre-test score of 57.3% and ended with an average score of 71.4%. The 2020 cohort started with 54.3% of the correct answers on the concept inventory and ended the semester with 65.9% correct.

In examining the gains in the reaction subscale, students who were in-person showed gains of a bit more than a quarter of a standard deviation (0.28) more than their peers who switched to online-only instruction midway through the semester. Though the reliability of the RXN subscale is lower than the overall reliability of the concept inventory (as would be expected because of the small number of items), it is interesting to note that the effect sizes of gains overall by cohort were consistent with those in the reactions subscale scores.

Gains were significant and strong for both groups, but the in-person group showed more progress overall and within the reaction subscale than the midway-online group. This reduction in effect size with a move away from hands-on IBLAs mirrors what was found for each of the non-hands-on activity approaches by previous work,[10–13; 19] where the demonstrations, simulations, demonstrations of simulations, and thought-experiments were all found to have significant yet smaller impact than the hands-on implementation.

| TABLE 2  
Concept Inventory Results |
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison of Means</strong></td>
</tr>
<tr>
<td><strong>Analysis of Gains within Cohort</strong></td>
</tr>
<tr>
<td>Subscale</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Entropy (ENT)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Reversibility (REV)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>U vs. H</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Equilibrium vs Steady State (EQ vs SS)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Reactions (RXN)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Overall</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*p ≤ .05; **p ≤ .01; ns = not significant (p>.05). Subscales in italics were conducted in the same manner for both cohorts.
Viewing statistics for each video are summarized in Table 3. Analogous statistics for Sp19’s in-person laboratories are unavailable but may be inferred from typical class attendance, which was over 90% (0-1 students typically absent on any given day). For the videos, time spent viewing the longer video was longer than for the shorter videos, but barely half of the viewers watched to the end of each video in one sitting. This is in contrast to hands-on experiments where it is nearly unheard of for a student to leave before the end of the experiment. Viewing statistics also indicate that over 95% of the views were associated with usernames in the university’s domain. This fact, taken with the number of views being at least 140% the size of the class enrollment, suggests that some students viewed the experiment multiple times, possibly only reviewing key elements on the re-watch.

It should be noted that while the novelty of the video format may have had a positive effect on learning, the overall stress of the abrupt change in instructional mode, combined with health, family, and societal concerns for many students, probably had a negative overall impact on student learning, and student performance in thermodynamics was likely affected in the same ways as their learning in other courses. As the post-test is taken at the end of the semester after the Sp20 students were sent off campus, it is quite possible this stress impacted performance for all sub-scales, not just the RXN sub-scale, as is suggested by the consistently lower post-test performance across sub-scales relative to the Sp19 cohort. It is possible that even when students complete simulations in class, the classroom environment itself is conducive to making greater gains because of the social nature of learning — the community aspect of everyone doing the same thing in the same place at the same time and the ability to interact and ask questions, etc. in real time during the simulations (see, e.g., ). The fact that both groups did so well may be partially attributable to the fact that most of the simulations that were used have been in place for more than a decade and have been well-tested empirically and experientially. Keeping in mind the small number of students in this study, the results provide preliminary support to the assertion that video is an acceptable mode of emergency replacement for hands-on conceptual laboratories, while suggesting hands-on experiments are preferable when possible.

### Table 3

<table>
<thead>
<tr>
<th>Video</th>
<th>Concept Area</th>
<th>Average % Viewed</th>
<th>Number of Views Before Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot Pot</td>
<td>Eq vs SS</td>
<td>63%</td>
<td>46</td>
</tr>
<tr>
<td>Explosive Reactions?</td>
<td>RXN</td>
<td>70.7%</td>
<td>41</td>
</tr>
<tr>
<td>Volcano</td>
<td>RXN</td>
<td>68.7%</td>
<td>34</td>
</tr>
</tbody>
</table>

**Application**

We would like to also offer guidance on experimental videos and the practices we used in developing ours. While we have not tested each of these recommendations, each was implemented in our work and so may be instrumental to the outcome. Based on literature and the work presented here, we continue to see that prediction and reflection questions, completed before and after an IBLA in any format (hands-on, demonstration, video, etc.), appear to be critical to learning and retaining the conceptual material. Instructors seeking inspiration for good conceptual questions will find a number of them in the AIChE Concept Warehouse and also at LearnChemE.com. The videos referenced in this study are approximately 1-2 minutes long to help maintain student attention and are available through this playlist: https://youtube.com/playlist?list=PLlF7UE2BwZ1k91VV4wzvkTlsZAKgSF_MK. When conducted hands-on, these experiments tend to take ~10 minutes each, due to set-up, clean-up, and other factors. The videos show the experiments at close range, in some cases nearer than students could safely be when conducting the experiments themselves. The experiments are also narrated by the instructor, with captioning for accessibility. It has been argued that the cognitive load from hands-on laboratories sometimes gets in the way of students’ conceptual learning within those spaces. For example, students may be so focused on writing down the next measurement or hitting the next button that they do not have time to reflect upon what the measurements mean. It seems likely that a video that is short and focuses nearly exclusively on the experiment and its outcome could lead to superior conceptual learning outcomes relative to a more realistic video that includes the entirety of the laboratory experience by removing details that distract from the conceptual core.

During in-person instruction, it is relatively easy for an instructor to confirm students have completed their predictions prior to engaging in the experiment. In remote instruction, it is helpful to leverage the capabilities of the LMS to encourage students to complete the assignment in the expected order. For example, through using an embedded quiz, students could be required to answer the prediction question prior to gaining access to the video link. Students could also be encouraged to come prepared to class or lab by a low-stakes, one-question quiz on the video material at the very start of class.

When replacing a laboratory with emergency remote instruction, it’s important for the instructor to decide which educational outcomes associated with a previously hands-on experience are most important to keep. Feisel and Rosa listed 13 canonical laboratory outcomes, most of which are possible to address through remote instruction although it is challenging for a single remote experience to address all of them. The experiments described here were designed to support conceptual learning as the main educational outcome, and therefore short videos that focus on the outcome of the
experiment seem to be an adequate replacement. Such videos are unlikely to be as successful at replacing other outcomes, such as data analysis, or gaining the knowledge of how to select and operate equipment.

Finally, it is vital to keep broad accessibility in mind when constructing materials for remote instruction. The written prediction and reflection questions may be shared online or in a physical handout that can be given or mailed to students. We recommend that the video make use of labels and clarifying audio so students know what they are seeing and what they are looking for. It is also helpful to include written descriptions and descriptive audio as an aid to visually-impaired learners. Captioning the video and providing students with access to the script are also important for learners who are working in a second language or have auditory challenges. The authors acknowledge that this work assumes students have access to internet speeds sufficient to stream video and a device (computer, phone, or tablet) capable of showing that video. High-definition video is attractive to capture and shows detail well, but adequate rendering may surpass the bandwidth video. High-definition video is attractive to capture and shows detail well, but adequate rendering may surpass the bandwidth video.

ACKNOWLEDGMENT

Hands-on thermodynamics IBLAs and CIET were developed with the support of the National Science Foundation, 0717536 and 0442234.

REFERENCES


Vol. 56, No. 1, Winter 2022