# Understanding Student Use of Worked Examples to Support Problem Solving via Real-Time Measurements of Problem-Solving Events

**Michael Crawford and Edward Berger**

**Purdue University, West Lafayette IN USA (crawfo20@purdue.edu, bergere@purdue.edu)**

***Abstract***: *Worked-example videos are ubiquitous and freely-available to students, and they are often used to support problem solving. The research suggests that learning by worked examples can be both effective and efficient for students. However, very few studies have examined the real-time problem-solving processes used by students. We conducted laboratory experiments in which participants solved a problem on paper, with access to a worked example video described as "potentially helpful". We collected real-time records of student problem solving events via a Livescribe system, video usage via an eye gaze capture system, and student thought processes via an audio recorded think-aloud protocol. Although few of their strategies for worked example use were systematic, many students nonetheless completed the problem with reasonable success. Worked-example video usage is a complex function of student sophistication and confidence as a learner, comfort with the subject matter, and individual goals about task performance and efficiency.*

## Introduction

### Multimedia Learning

The past two decades have seen enormous strides in all manner of digital technologies, as well as the personalization of learning based upon many of those technologies. The ubiquity of multimedia content through YouTube and other channels provides students with countless free resources for learning. The scholarship around multimedia learning has also developed, led by Mayer (Mayer, 2009; Mayer & Moreno, 2003) and others (Berger & Krousgrill, 2012; Brunken, Plass, Leutner, & Brünken, 2003) who developed important frameworks to help us understand the appropriate role of multimedia resources for learning, as well as how to construct multimedia resources effectively. The challenge today is one of scale and strategy: given a specific learning task, how do students (i) select an appropriate, useful video from the massive library of available videos to help them with that task, (ii) how exactly do they use that video while executing the task, and (iii) in what ways can their professors coach them in optimizing their effort and maximizing their performance on that task? These are critical questions, and this paper focuses squarely on the second.

### Research Question

Our focus on the ways in which students use worked-example videos to support their problem solving generates our **research question**: to what extent are there detectable patterns to student usage of a worked-example video during problem solving? The working hypothesis is that there exist a small number of usage patterns for worked-example videos, and that our approach to the experiment will allow us to observe those patterns.

## Theoretical Frameworks

### Worked-Examples and Learning Trajectory

The worked-example effect has been widely studied for teaching and training in many disciplines (Paas & van Gog, 2006; Rourke & Sweller, 2009; Smith, Mestre, & Ross, 2010;

Sweller & Cooper, 1985; Ward & Sweller, 1990), as well as a particular instantiation of multimedia learning (Moreno, 2006; Moreno, Reisslein, & Delgoda, 2006; Moreno, Reisslein, & Ozogul, 2009). The underlying construct of worked-example research is cognitive load theory (CLT), which posits that humans have finite cognitive capacity that is the summation of three dimensions: germane (related to schema automation and commitment to long-term memory), intrinsic (related to the inherent difficulty of the material to be learned), and extraneous (related to the way the material is presented). The idea behind worked examples is that if they are well designed, they can help learners optimize cognitive load by minimizing the extraneous component and diverting available cognitive capacity to the germane load of schema formation/automation. Video-based worked examples provide the added benefit of being replayable, sharable, portable, and device-independent (Berger & Krousgrill, 2012).

However, worked-examples as learning tools have diminishing returns as the learner becomes more sophisticated, and this is called the expertise-reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Salden, Aleven, Schwonke, & Renkl, 2009). Broadly speaking, this phenomena suggests that learners who are new to a subject can benefit substantially from watching experts solve problems related to that subject (i.e., watching well-constructed video worked examples). However, as students become more competent problem solvers, the relative utility of worked examples decreases, and their learning improves more readily when they simply practice solving problems themselves. This notion of a learning trajectory, and the placement of worked-examples along it, helps us understand the research as well as it implications for engineering education practice.

### Illusion of Explanatory Depth

To understand the implications of this research for engineering education practice, the illusion of explanatory depth (IOED) provides a useful framework. The seminal IOED study (Rozenblit & Keil, 2002), through a clever series of experiments, illuminated the severely limited ability of test subjects to accurately estimate their own level of knowledge about reasonably complicated topics, such as devices (ex.: a toaster) or natural phenomena (ex.: lightening or earthquakes). The typical trend was for a test subject to wildly over-estimate their level of knowledge until confronted with an assessment and/or an expert explanation of the topic, at which point a more realistic (although still inaccurate) self-rating of knowledge emerged. This systematic prediction inaccuracy was not valid for fact-driven topics, such as narratives of a movie or world capitals; subjects generally had a good sense of their own knowledge of, say, the capital of Colombia. But for complex, nuanced topics—and engineering problem solving falls into this category—the typical behavior of test subjects was to provide very poor predictions of their actual knowledge until confronted with an assessment. *The essence of IOED is that people often believe that they understand or can do that which they have merely seen*, and the potential relationship to worked-example videos is clear. The implication in this research, expanded upon later in the paper, involves the way in which we coach students to use worked-example videos to support problem solving or test preparation so that they can avoid the pitfalls of IOED.

## Methods

### Experimental Apparatus and Data Collection

We collected data from 24 experiment participants, all undergraduate engineering students, and have analyzed the data in detail for 10 of them so far. The full experimental setup has been described in detail elsewhere (Berger & Wilson, 2016), but several of the salient details are summarized here. Figure 1 shows a screenshot of the outcome of the experiment, including the three data collection modalities used. Each participant, after having been consented for the purposes of this research according to our approved IRB protocol, completes a single engineering dynamics problem on paper using a Livescribe notebook (which captures a real-time record of everything they write). The problem focused on particle kinetics and had three sub-parts: (a) requiring a work-energy formulation, (b) utilizing

Newton's law, and (c) using a different work-energy equation from part (a). Simultaneously, participants have access to a worked-example video of an expert solution to a dynamics problem (duration: 04:23) that shares several, but not all, features of the problem they are trying to solve. An eye gaze capture system plus video screen capture software records, in real time, the participant's interactions with the video including play, pause, stop, and gaze events. Finally, participants are asked to think aloud and narrate their thought process throughout the experiment; this audio is capture via the Livescribe pen. These data streams are merged into a single time-synchronized video (see the screenshot of Figure 1).

Participants also completed a pre-survey about their comfort with technology for learning, and a post-survey based upon the NASA-TLX workload measurement (Hart & Staveland, 1988). Participants also consented to allow us to access their academic transcript, so we know their grade in the dynamics course and all its pre-requisites.

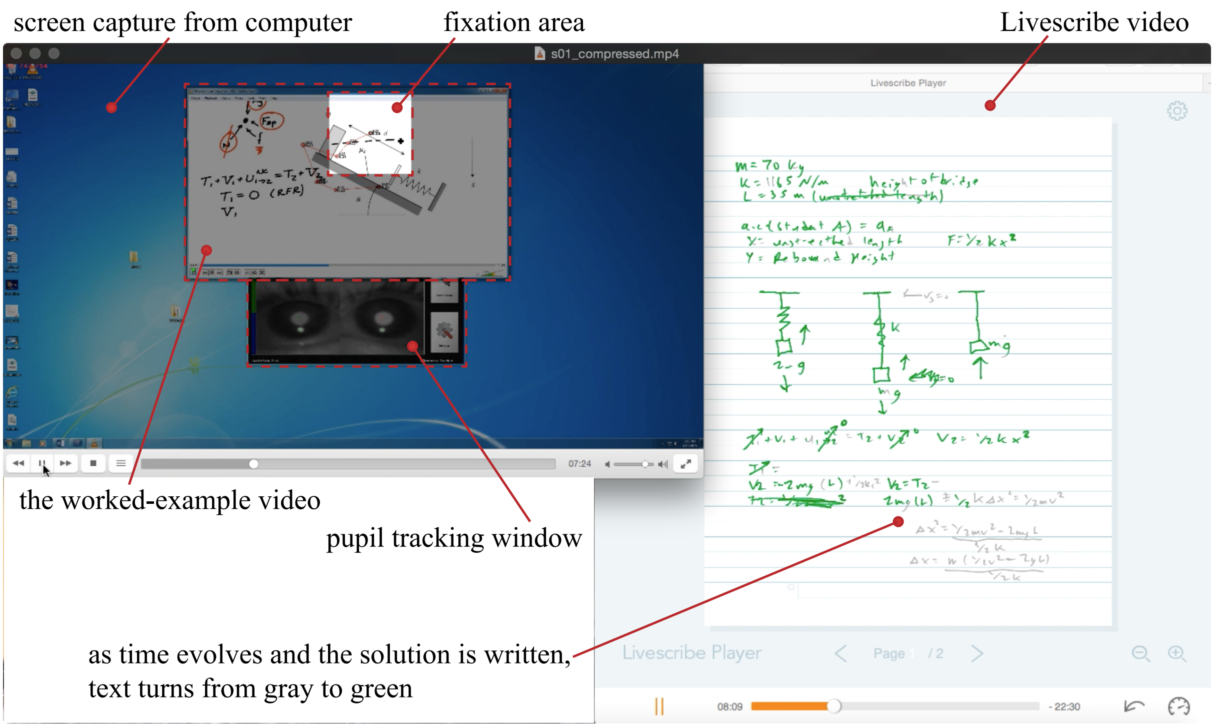


**Figure 1: Still frame from experimental composite video showing both eye gaze capture panels and Livescribe video [from** (Berger & Wilson, 2016)**].**

## Data Analysis

The resulting video recording of each experiment was transcribed by a dynamics content expert to capture the following features: (i) the think-aloud narration from the participant, (ii) the specific solution events for the problem (ex.: drawing a free body diagram), and (iii) interaction events with the video. This experimental transcript is non-trivial to compose, and particularly solution-specific events really do require the coder to be a content expert. The transcripts created and coded for this research were processed by the authors (a faculty member who teaches the dynamics course, and an advanced undergraduate mechanical engineering honors student).

Each transcript was then coded according to a previously-developed code structure, with themes and sub-themes related to solution events, video events, and affect events. Affect events were generally audible expressions by participants that expressed confidence, frustration, confusion, or any other emotion throughout the experiment. This coding structure has also been presented in detail elsewhere (Berger & Wilson, 2016). We highlight here the category of 'forward reasoning' events, which capture participants advancing their solution in one of three ways: (i) using verbal or written information, (ii) using sketches or drawings, or

(iii) using algebraic operations. Forward reasoning is one of the key facets of the solution approach and one important metric we consider characteristics when comparing participants.

After coding, the data were processed to determine code counts, frequencies, time durations, sequencing, and other metrics related to the solution process. An example outcome is the coding stripe output of Figure 2, which shows the categories of events, their duration, and their sequencing. The coding stripe figure reports on each of the broad themes observed in the data; from top to bottom on the figure, they are: (i) which part of the problem they were solving, (ii) their engagement with the video, (iii) their forward reasoning events, (iv) their other solution events, and (v) their affect events. On the right side of the figure, the frequency (F) and duration (D, in minutes) of each event is shown. Frequency and duration for each event was calculated and summarized across subjects using descriptive statistics.

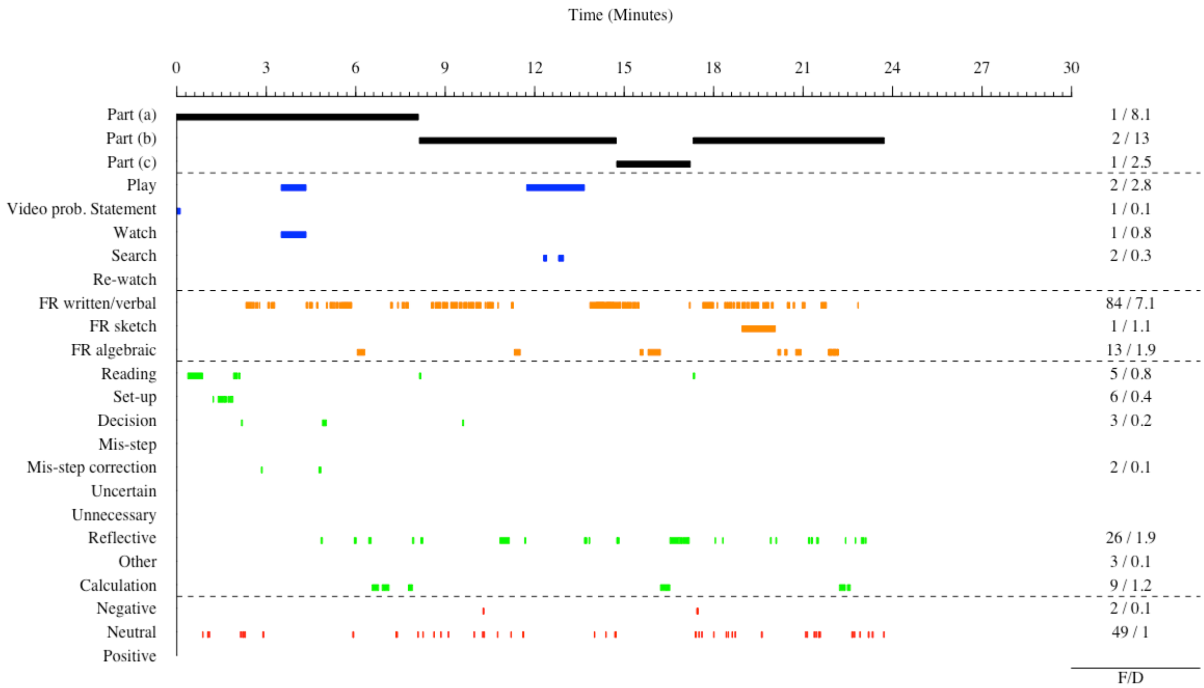


**Figure 2: Coding stripe output of experimental coding for subject #7. The coding themes are listed on the left side, while the frequency (F, number of instances) and duration (D, in minutes) for each event/theme are on the right side.**

# Findings

## Data Analysis Results

We can use the coding stripe figures, such as Figure 2, as a way to understand the temporal patterns used by participants during the problem solving experiment. Figure 3 shows a coding stripe comparison for five test subjects across a subset of the themes described above and shown in Figure 2. These sub-themes were identified in our exploratory data analysis as capturing much of the variation across participants. In Figure 3, we focus on time devoted to each part of the problem (black stripes), time devoted to playing and watching the video (blue stripes), and time devoted to the three categories of forward reasoning (orange stripes).

The distribution of time spent on each part of the problem generally favors part (a), likely because part (a) contains many of the start-up portions of the problem (such as drawing a free body diagram, defining a coordinate system, and understanding/modeling the overall dynamics of the problem). Many of these achievements in part (a) can be directly applied to parts (b) and (c), meaning that those parts of the problem are likely to take less time to complete.

The variation in video usage is instructive in two respects. First, there was wide variation in the amount of time each participant played the video during the experiment. Moreover, it is clear that video 'play' time and 'watch' time are not necessarily the same—or even similar. During the experiment, we saw many cases in which students played the video in the background, listening to the audio without actually watching the video, while they were thinking about the problem. It is conspicuous that generally the forward reasoning events took place when the video was not playing or being watched.

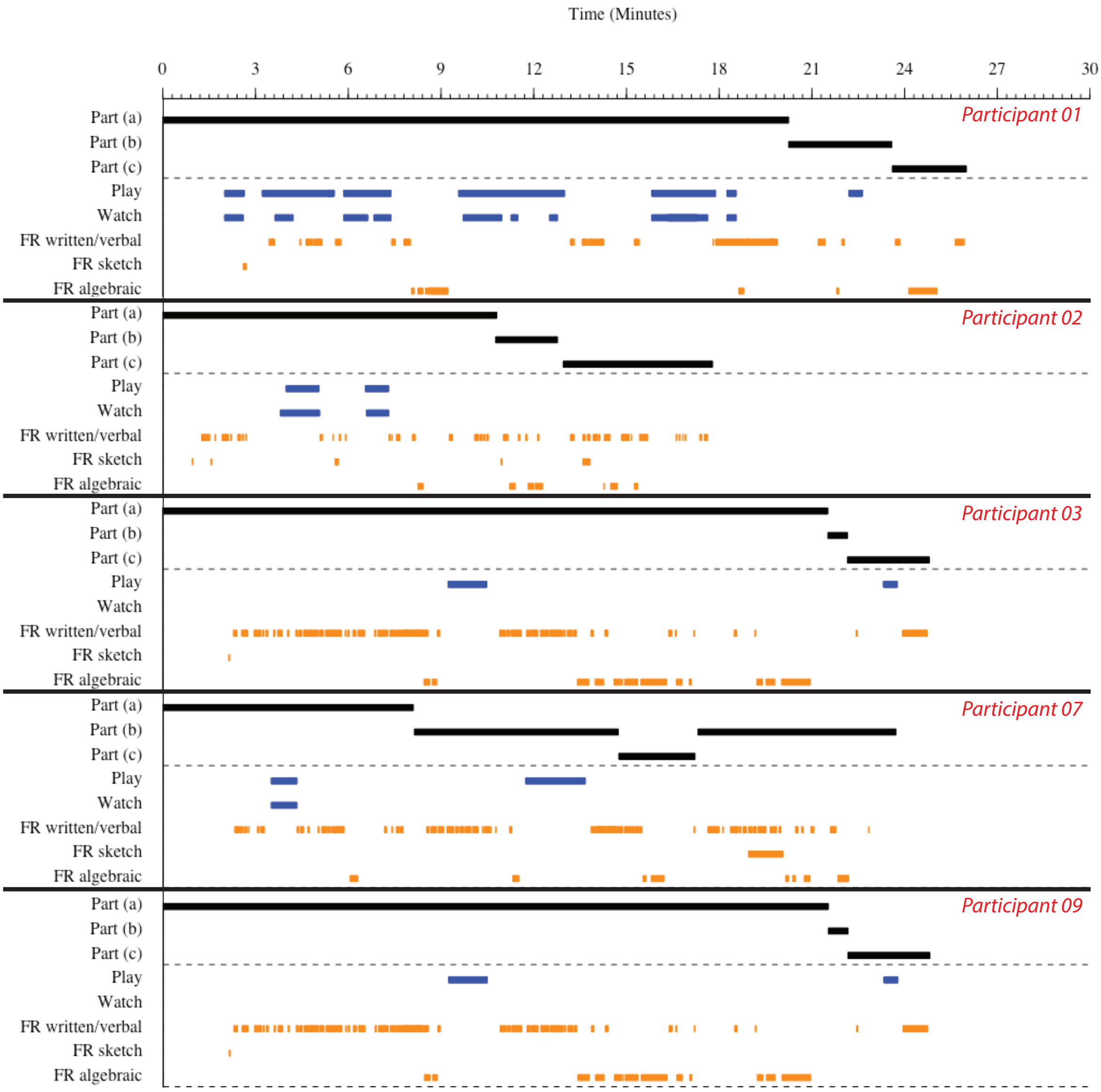


**Figure 3: Code stripe composite showing several of the most important themes for five different participants. Parts (a), (b), and (c) correspond to the three parts of the problem. 'Play' and 'Watch' refer to participant interaction with the video. The 'FR' categories describe the three types of forward reasoning.**

Table 1 summarizes some of the key time metrics across all 10 participants in this phase of our study. Participants generally used almost 2/3 of the total experiment time to complete part (a) of the problem, for the reasons described above. As observed from the coding stripe figure, there is a dramatic difference between the play and watch time for the video, and the ratio of the percentage of time spent playing to time spent watching was 19/7 = 2.7. Although not shown on the table, the percentage of time participants spent searching for content within the video was very small, just 0.6% of the total time for the experiment. Table 1 also makes it clear that these participants generally favor written/verbal modes of forward reasoning, rather than sketching or mathematical (algebraic) approaches.

**Table 1: Time duration summary for specific codes across all 10 participants. Metrics are listed as mean and standard deviation of time (in minutes), and as a percentage of the total time taken to complete the problem.**

| *Code* | *Time (minutes; mean/SD)* | *% of total time* |
|---|---|---|
| Part (a) | 12.1/5.6 | 62.0 |
| Part (b) | 3.7/3.8 | 18.0 |
| Part (c) | 4.0/3.2 | 20.0 |
| Video play | 3.5/3.7 | 19.0 |
| Video watch | 1.5/2.2 | 7.0 |
| FR—written/verbal | 5.9/1.9 | 32.3 |
| FR—sketch/drawing | 0.7/0.8 | 4.2 |
| FR--algebraic | 1.7/1.1 | 8.1 |

Participant verbal accounts of their work generally did not give the impression of intentional use of the video. We often heard speculative comments such as ("[the video honestly doesn't look that helpful" (Participant 3), or "video is kind of useful" (Participant 1). Several participants simply started the video at its beginning and watched the video to assess its content and potential usefulness for their problem solution. We did observe several instances of pattern matching, in which equations from the video were mimicked and used in the participant's own solution to the experimental problem.

## Discussion

Two general patterns emerged from our research results.

1. <u>Students did not know how to use the video efficiently</u>. The ratio of 'play' time to 'watch' time was a surprising 2.7, clearly suggesting that many students played the video while they were working without paying very close attention to it. Moreover, very little time was spent on intentional searching within the video for any participant (and only about 1 minute of searching in the aggregate, across all 10 participants), suggesting that they: (i) searched highly efficiently, (ii) did not know what to search for, or (iii) did not employ a study strategy that included searching videos for specific pieces of information (i.e., their 'usual' approach was to watch large chunks of video). Based upon our observations during the experiment, option (i), highly efficient searching, simply did not occur and we discount this as a possible explanation for the very low search time. More likely, and consistent with the high play/watch ratio, is that *participants did not possess the required insights of abilities to enable efficient search within the video or recognize portions of the video that could have helped them solve the experimental problem more efficiently or with higher performance*. Their choice to play the video, often in large chunks, suggests that they were attempting to engage in pattern matching with the video's overall problem solving approach, rather than picking out the specific features of the video that would allow them to advance their own solution.
2. <u>Most video-inspired forward reasoning seems to be pattern matching</u>. The other observation is that forward reasoning and video watch were rarely synchronous in our data. Video watching and forward reasoning were often sequential, but our observations from each experiment as well as the audio of the experiment do not suggest that

watching the video directly led to a ‘breakthrough’ or induced specific conceptual understanding by the participant. The sequence of video watching and then forward reasoning was more likely related to pattern matching (say, using the equation used in the video) than it was to any improvement in conceptual understanding by the participant. *The conclusion here is that for these participants, video usage did not induce improved conceptual understanding, so much as it triggered a pattern matching behaviour.* This observation highlights differences between the current study (about using worked-example videos as an aid to problem solving) and past worked-example research (which often focused on teaching and training via worked examples). It appears that the sequence of video playing or watching and forward reasoning evident in the code stripes (Figures 2 and 3) are more likely the result of an ad-hoc approach to solving the experimental problem. Participant usage of the video seems to have been more hopeful than strategic, in which they used the video to identify a solution approach and equations that they could use in the experimental problem, rather than attempting to improve their fundamental understanding of dynamics.

Taken together, we believe both of these patterns have serious consequences in light of the IOED. The evidence presented here suggests that *students work with videos for the purpose of ‘doing’ rather than ‘learning’*. They do not seem to view the video as an expert source from which they can learn, so much as they see it as a resource to help them achieve an objective (completing an assignment). The danger in this strategy is obvious: students can convince themselves that they understand a topic by simply matching a pattern observed in a video to solve a problem, when in fact their understanding of the topic might be highly superficial. This is the essence of the IOED and has important implications for how we coach students to use worked-example videos, as described later.

### Limitations

There are two important limitations to this study. First, we recognize the small sample size in the data so far—just 10 participants. While the experiment itself takes about 45 minutes, including the pre- and post-surveys, the data analysis of all the video evidence, transcript generation, and coding is an extremely time-intensive process. We continue to analyze the experimental data and build our sample size.

The other limitation relates to the measurement of watch time. We suspect that our eye gaze equipment is producing a lower-bound estimate of the actual watch time, for the following reasons. ‘Watch’ time was only tabulated when the participant’s pupils were visible in the pupil window of Figure 1, meaning that their head position and angle were aligned with the initial set-up and calibration of the system at the beginning of the experiment. In the presence of misalignment in head position, the eye gaze system still measures a gaze location, but we believe that location estimate to be in error based upon the geometric differences in participant position from the calibration condition. As a result, given this strict definition of ‘watch’ time, we believe the actual watch time is larger than that reported here (resulting in an actual play/watch ratio of less than 2.7). However, based upon our observations during the experiment, we do not believe that the reported watch times were dramatically different from the actual watch times. Nonetheless, addressing this experimental set-up issue will be an important refinement to future experiments.

## Implications for Engineering Education Practice

### Student Coaching

One important implication of this research is that students do not seem to make optimal use of the videos, in two respects. First, their pattern matching behaviour is not desirable from either a pedagogical or IOED perspective, and we would rather they take insight and meaning from the video. Second, their lack of efficient search processes suggests that they cannot articulate or recognize that which they do not know. That is, they do not know what they are looking for in the video, and perhaps they do not recognize it when they see it. Taken together, these two observations suggest that we must do a better job of explicitly coaching students on how to make the best use of the video resources.

The current generation of college students has grown up with YouTube, and it is natural to assume that their sophistication in using videos is very high. However, using worked-example videos for problem solving requires two important actions: (i) select the appropriate video that has the strongest alignment with the problem to be solved, and (ii) find the specific elements/concepts/processes within the selected video that most effectively enlighten the viewer about proper solution procedures. Coaching students about this process could be invaluable in both their performance (they earn a better score) and their efficiency (they know how to quickly identify useful information, and therefore can complete the assignment more quickly). The evidence presented here strongly suggests that students would benefit from more intentional coaching about how to use worked-example videos.

### Video Creation

This research also holds value for video producers. Given what we know about student approaches to pattern matching, perhaps video creators can explicitly indicate the contents of the video using more extensive metadata, chapter markers, or other guideposts to help students evaluate the utility of a given video. Other options include explicit links to existing taxonomies or tools such as concept inventories for the video contents. Moreover, the voiceover within the video could include more than the simple mechanics of the solution (first we do this, then we do that) and instead include expert insights about why the video proceeds as it does. For instance, the video creator could intentionally describe the thought process behind a particular approach: “I’m using a work-energy approach because I’m attempting to solve for a change in position rather than a change in time…” These expert insights would certainly enlighten student users of the videos, and also may shift their mindset from one of pattern matching to a more intentional approach of video usage for learning in the service of problem solving.

## Conclusions

This research reports the results of 10 experiments that help us understand student usage of worked-example videos to support their problem solving. Our careful transcripting and coding of the experimental data suggest that students engage in strong pattern matching efforts rather than a sophisticated search approach when using worked-example videos. The results also suggest that video watching does not trigger improvements in ‘understanding’ so much as it triggers pattern matching and mimicry of the processes used in the video. The implication for engineering education is that explicit coaching for students about how to most effectively use worked-example videos seems beneficial, as do improved video creation guidelines that include more detailed metadata and fully articulated expert insights. Our on-going data analysis continues to enhance our understanding of student use of worked examples, with further advances to be reported in future publications.

## References

Berger, E. J., & Krousgrill, C. M. (2012). HigherEd 2.0: Web 2.0 in Higher Education. In I. Deliyannis (Ed.), *Interactive Multimedia* (pp. 67–92). Rijeka, Croatia: InTech Europe.

Berger, E. J., & Wilson, M. (2016). A Laboratory Study of Student Usage of Worked-example Videos to Support Problem Solving. In *Proceedings of the 123rd American Society for Engineering Education Annual Conference and Exposition* (p. 17 pages/Paper ID #16199). Seattle, Washington.

Brunken, R., Plass, J. L., Leutner, D., & Brünken, R. (2003). Direct Measurement of Cognitive Load in Multimedia Learning Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 53–61. http://doi.org/10.1207/S15326985EP3801

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results and Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental*

*Workload*. Amsterdam: North Holland Press.
Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 23–31. http://doi.org/10.1207/S15326985EP3801
Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). New York: Cambridge University Press.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52.
Moreno, R. (2006). When worked examples don't work: Is cognitive load theory at an Impasse? *Learning and Instruction*, *16*(2), 170–181. http://doi.org/10.1016/j.learninstruc.2006.02.006
Moreno, R., Reisslein, M., & Delgoda, G. (2006). Toward a fundamental understanding of worked example instruction: Impact of means-ends practice, backward/forward fading, and adaptivity. *Proceedings. Frontiers in Education. 36th Annual Conference*, 5–10. http://doi.org/10.1109/FIE.2006.322285
Moreno, R., Reisslein, M., & Ozogul, G. (2009). Optimizing Worked-Example Instruction in Electrical Engineering: The Role of Fading and Feedback during Problem-Solving Practice. *Journal of Engineering Education*, *98*(1), 83–92.
Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87–91. http://doi.org/10.1016/j.learninstruc.2006.02.004
Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, *19*(2), 185–199. http://doi.org/10.1016/j.learninstruc.2008.03.006
Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: an illusion of explanatory depth. *Cognitive Science*, *26*(5), 521–562. http://doi.org/10.1207/s15516709cog2605_1
Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2009). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*(3), 289–307. http://doi.org/10.1007/s11251-009-9107-8
Smith, A. D., Mestre, J. P., & Ross, B. H. (2010). Eye-gaze patterns as students study worked-out examples in mechanics. *Physical Review Special Topics - Physics Education Research*, *6*(2), 1–9. http://doi.org/10.1103/PhysRevSTPER.6.020118
Sweller, J., & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*, *2*(1), 59–89. http://doi.org/10.1207/s1532690xci0201
Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1–39. http://doi.org/10.1207/s1532690xci0701

## Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant DUE-1525671. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## Copyright Statement