Understanding the Effect of In-Video Prompting on Learners and Instructors

Hyungyu Shin[1]   **Eun-Young Ko**[1]   **Joseph Jay Williams**[2]   **Juho Kim**[1]

[1]KAIST, Daejeon, Republic of Korea
[2]National University of Singapore, Singapore, Singapore
{hyungyu.sh, eunyoungko, juhokim}@kaist.ac.kr   williams@comp.nus.edu.sg

ABSTRACT

Online instructional videos are ubiquitous, but it is difficult for instructors to gauge learners' experience and their level of comprehension or confusion regarding the lecture video. Moreover, learners watching the videos may become disengaged or fail to reflect and construct their own understanding. This paper explores instructor and learner perceptions of *in-video prompting* where learners answer reflective questions while watching videos. We conducted two studies with crowd workers to understand the effect of prompting in general, and the effect of different prompting strategies on both learners and instructors. Results show that some learners found prompts to be useful checkpoints for reflection, while others found them distracting. Instructors reported the collected responses to be generally more specific than what they have usually collected. Also, different prompting strategies had different effects on the learning experience and the usefulness of responses as feedback.

ACM Classification Keywords

K.3.1. Computer Uses in Education

Author Keywords

Online Education; In-Video Prompting; MOOC; Reflection; Feedback; Learners; Instructors.

INTRODUCTION

In online learning environments, an important task for instructors is to inspect learners' level of comprehension and learning experience regarding their lecture videos. Such inspection enables instructors to gain insights into what to teach and how to teach in future instruction. Online learning platforms such as Coursera and edX collect data from multiple sources and provide dashboard interfaces for instructors to support this task. Available data streams include video clickstream logs, responses to in-video quizzes, submissions for assignments and exams, platform interaction data, discussion forum posts, and course reviews.
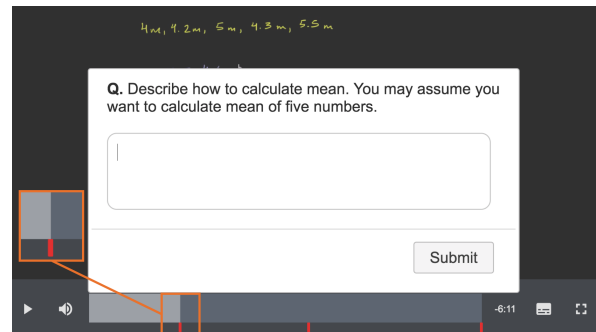
Figure 1: An example of in-video prompting supported by a video learning interface. Once the playhead reaches the red bar, the prompt appears. This example asks a question about the learner's comprehension on a specific part of the video. While common, it's not the only possible prompting type. This paper explores the design space of in-video prompting, and presents how learners perceive the general idea of in-video prompting and different prompting designs.

Research shows that learner-generated data such as artifacts in discussion forums and learner surveys are important in monitoring online courses [20] and preparing future iterations of the course [27]. Discussion forums and review websites are popular communication channels where learners share their comprehension and evaluation of the course. Through discussion forums, instructors can check how learners understand concepts and correct misconceptions by intervening in discussions among learners [15]. Review websites address various dimensions of a course, such as the level of difficulty, quality of contents, and instruction delivery. From these channels, instructors can assess how learners perceive their material and course design, which could serve as useful feedback.

However, existing channels for collecting learners' responses and input are insufficient as a source of feedback. Posts on discussion forums could give insights into the level of comprehension, but only a small portion of learners participate in the discussion forums [3]. Posts on review websites capture subjective learning experiences, but their granularity is at the course level, which is not specific enough for instructors to identify which specific parts of the lecture need to be improved.

To address these challenges, we investigate **in-video prompting** as a channel for collecting learners' feedback on lecture

Table 1: The design space of in-video prompting questions, segmented by comprehension-experience orientation and the level of specificity. Comprehension-centered/Experience-centered questions ask about learners' comprehension and learning experience. General/Specific questions determine whether the questions refer to lecture content or not. Table 1a shows example questions for each prompting strategy, and Table 1b shows sample responses for each prompting condition.

	Comprehension-centered	Experience-centered
General	Describe what you have learned so far.	Describe something unsatisfying about the lecture so far.
Specific	Describe how to calculate the standard deviation. You may assume you want to calculate the standard deviation of five numbers.	Describe something unsatisfying about calculating the standard deviation. You can consider how clear the explanation was, how fast the explanation was, and what information was missing.

(a) Sample questions for each prompting condition.

	Comprehension-centered	Experience-centered
General	Population standard deviation and how it is calculated and what it means if it is a larger number.	The instructor's handwriting is a little bad.
Specific	The standard deviation is calculated by taking the square root of the variance.	Explanation of change from squared units to units was slightly too fast.

(b) Sample responses to each question presented in Table 1a, submitted by crowd workers in our study.

videos while minimizing disruption in learners' experience. In-video prompting presents reflective questions to learners *while* they are watching the video, in order to get specific comments on their level of comprehension or learning experience. Figure 1 shows an example video learning interface equipped with in-video prompting. Unique properties of in-video prompting are that (1) all learners who watch the video encounter the prompts, which is likely to yield a high response rate, (2) prompting at an inopportune time might distract learners from learning, and (3) since the prompts are given in the middle of learners' video watching session, they can ask specific questions to learners about the material just covered in the video.

Little research has investigated the effect of in-video prompting and its design space. A common type of in-video prompting is in-video quizzes [13], which are commonly multiple choice questions or short-answer questions that pop up during playback to maintain engagement and check understanding. But the types of possible in-video prompts are not limited to quizzes. For example, they could allow a more detailed insight into learners' interpretation of video content. Also, carefully designed prompts could enhance learning, as research shows that reflective prompts improve learning outcomes [5, 24].

The design of questions used for prompting is important in determining the type and quality of comments that instructors collect and students' learning experience. Depending on what goal the instructor wishes to achieve with in-video prompting, questions could focus on either revealing learners' level of comprehension, or understanding their subjective learning experience. As the learning experience is affected

by the questions, we also need to understand how learners perceive in-video prompting. To understand the effect of in-video prompting on both learners and instructors, we pose the following research questions:

- RQ1. What are learners' perceptions of in-video prompting?

- RQ2. How do different in-video prompting strategies affect the learning experience?

- RQ3. How useful are learner responses to in-video prompts as feedback to instructors?

We focus on two dimensions in exploring the design space of in-video prompting: the type of information collected from learners, and the specificity of the question. Table 1a illustrates the combinations of these two dimensions (the comprehension-experience orientation and the level of specificity) with example questions. The comprehension-experience orientation represents which information learners need to submit: comprehension-centered questions ask about the level of comprehension, while experience-centered questions ask about the learning experience. The level of specificity represents whether the questions refer to lecture content or not. General questions prompt for reflection on the lecture content without referencing specific context, whereas specific questions ask about the lecture content in detail. Table 1b shows examples of learner responses for each prompting questions.

To answer the three research questions, we conducted a series of studies and interviews. To understand the learners' perspective, we conducted two studies with crowd workers to explore both the effect of prompting in general (RQ1) and the

effect of different prompting strategies (RQ2). We collected open-ended responses about the pros and cons of prompting from learners and found evidence that learners generally prefer comprehension-centered prompting to experience-centered prompting. We conducted interviews with instructors and instructional designers to understand the usefulness of the collected learner responses as feedback on how the learners were doing, and if the course iterations needed change (RQ3). It turns out that collected responses are generally more specific than what they have usually collected. Instructors found responses to experience-centered questions provided more actionable feedback than those from comprehension-centered questions. Instructional designers provided insights into designing better prompting strategies.

The contributions of this paper are as follows:

- Results from a study on learners' perception of in-video prompting, organized into a list of pros and cons.

- Results from a study on learners' perception of four different prompting strategies, which span general versus specific, and comprehension-centered versus experience-centered.

- Results from interviews with instructors and instructional designers about the usefulness of learner responses as feedback on lecture videos.

The rest of paper is organized as follows. First, we survey related work. Second, we present instructor perspectives on learner feedback on lectures. Third, we present two studies designed to understand how in-video prompting affects learners. Fourth, we report results from interviews with instructors and instructor designers to understand how learner responses to in-video prompting serve as feedback. Finally, we conclude with the discussion of the effect of in-video prompting.

RELATED WORK

In-video prompting involves promoting learner reflection and providing feedback to instructors. We briefly review prior research on these topics, such as methods for collecting feedback in both offline and online settings and the effect of reflective prompting on learners.

Utilizing Learners' Interaction Data

Existing online learning platforms have been collecting learners' interaction data in both passive and active ways. Clickstream and in-video dropout data are passively collected in that the data is naturally collected regardless of learners' intention. Although research has investigated presenting [11] and analyzing passive data [4, 26], the conclusions that can be drawn are limited as feedback because users' true intentions behind traces are unknown. Forum posts are actively collected in that learners explicitly give the data to the platforms. Agrawal et al. [1] and Wen et al. [22] have investigated analyzing discussion forum posts to get meaningful insights from them. The granularity of forum posts is at the course-level, which makes it hard for instructors to figure out which specific parts of the lecture need to be improved. Singh et al. [19] and Lee et al. [14] facilitate discussion among learners within a lecture video. Although instructors can get more specific learner comments,

it has limitation as feedback in that instructors cannot control the types of learner comments. In this paper, we explore in-video prompting, which yields specific and controlled learner comments, as a channel for collecting more useful feedback.

Providing Feedback on Online Lecture Videos

Vidcrit [17] supports asynchronous video review and sharing feedback on videos, but is designed for reviewers who indicate problems or offer suggestions, not for learners. Mudslide [9] attempts to collect spatially contextualized 'muddy' points in a video with specific explanations, which enables instructors to figure out the common points of confusion with reasons. We see Mudslide as a case of in-video prompting, which uses spatial anchoring at the end of the video as its prompt. In this work, we aim to understand the effect of in-video prompting broadly by exploring two specific design dimensions described in Table 1a.

Collecting Qualitative Comments in Offline Classroom

In a physical classroom, classroom assessment techniques (CAT) can be used to collect qualitative comments [8]. A technique described in CAT is called One-Minute paper, which asks questions to students at the end of the class. The questions include "What are the most important concepts you have learned today?", and "What are the most confusing points?". It not only provides students with a better learning experience [21] and higher scores on tests for some cases [7], but also enables instructors to improve teaching with formative evaluation. Our work seeks to explore the design space of in-video prompting and to help design effective online prompts that provide benefits to both learners and instructors like the One-Minute paper.

Effect of Reflective Prompting on Students

Many studies have demonstrated that students can learn more when they are prompted to explain the meaning of what they are learning. For example, prompting students to explain what they understand from biology texts enhances the accuracy of students' mental models about the circulatory system, arguably by helping students generate inferences and spot gaps in their understanding [5]. Williams et al. [24] suggest that explaining why a fact is true drives learners to discover underlying patterns or principles. On the other hand, there are many known cases where prompts to reflect do *not* enhance learning, and many more that are likely unreported. There are even cases where prompts to reflect can *hurt* learning, by causing learners to overgeneralize [25], ignore details [23], or rely on incorrect prior knowledge rather than observed facts [6]. These contradictory finding underscore the importance of exploring the design space of how prompts to reflect impact learners. In particular, it is important to understand the strengths and weaknesses of different *kinds* of prompting strategies, and how these are perceived by learners.

In the context of learning online video, platforms like Coursera allow instructors to insert in-video multiple choice quizzes. While there have been studies of how they affects learners' video navigation [13], there is less evidence about the causal impact on learning. In addition, in-video quizzes must be

designed for the specific content of a video, while work on re-flective prompting tends to use general prompts that aren't tied to specific content. There has been relatively little research on the effects of adding general prompts to online videos, as existing work focuses more on different formats of video presentation [16]. In addition, educational studies of prompting have focused more on learning outcomes than on learners' subjective experiences. And as most studies of prompting have been conducted in laboratory settings or physical classrooms, little is known about what instructors might learn by being able to rapidly view students' responses to prompts. In this work, we attempt to fill this gap in literature.

INSTRUCTOR PERSPECTIVES ON LEARNER FEEDBACK ON LECTURE VIDEOS

To better understand instructors' perspectives on learner feedback on lecture videos, we conducted interviews with three instructors on campus and a web-based survey with five MOOC instructors outside of campus. We aimed to understand (1) what learner feedback instructors collect in their current practice, and (2) how useful learner feedback is in improving instruction. Among the interviewees, one instructor had experience in teaching a MOOC, and two instructors had experience in the flipped classroom method. Each interview session took about an hour, and the expected completion time for the survey was 15 minutes. We summarize the main findings below.

Lack of learner comments. Instructors reported having few learner comments to work with in the first place. In the interview, the MOOC instructor said there were approximately 10 questions per each video. The flipped classroom instructors pointed out that they do not typically ask for feedback because students tend to passively consume the lecture video and they did not expect students to be willing to provide useful comments.

Lack of Specificity. Even in the context where learners leave comments in forums, instructors responded that the comments are not specific enough to understand what causes the confusion or problem. Two of the survey respondents expressed that they would like to receive comments specifically anchored to the lecture content, such as *"I would like more examples of this algorithm"*, and *"How do Japan and China name Japanese invasions of Korea described in slide 17?"*. One of the survey respondents expressed the need for feedback on her instruction delivery, such as the tone of speech and sentence length.

From the interviews and the survey, we confirmed that there are few learner comments, and the comments are generally not specific enough to be used as feedback. We anticipate that in-video prompting can address these challenges by actively asking questions while learners are watching lecture videos. However, it is hard for learners to provide a large amount of specific feedback. Questions that drive our investigation in in-video prompting include: How should we design questions in prompts to collect useful feedback? What is the effect of prompting on the learning experience? How distracting is it for learners to answer the questions while watching a video? How useful are the collected responses as feedback to instructors? To answer these questions, we conducted a series of studies.

INVESTIGATING IN-VIDEO PROMPTING

We aim to understand the effect of in-video prompting on both learners and instructors. Our investigation is organized as follows. We first present two studies designed to understand the effect of in-video prompting on learners. Then we consider the usefulness of collected responses as feedback from instructors' view as well as from instructional designers' view. Finally, we wrap up the studies by discussing the complexity of considering viewpoints of multiple stakeholders when designing prompting strategies.

We conducted two studies to understand the effect of in-video prompting on learners. The first study explored learners' qualitative experience of receiving in-video prompts. Learners watched videos with and without reflective prompts, and answered open-ended questions about their experience. The second study investigated how the *type* of prompt might influence learners' perceptions, comparing specific versus general prompts, and prompts focused on comprehension versus sharing one's experience with the instructor. The second study also collected quantitative measures of learners' experience.

Learners received prompts to reflect at the beginning, middle, and end of each video. Prompts at the beginning could prepare people for learning [18], prompts in the middle can maintain engagement with mid-video prompts, and prompts at the end help in review the material as a whole [10].

STUDY 1. LEARNER PERCEPTIONS OF PROMPTING

The objective of the first study was to gain a qualitative understanding of how learners perceive the addition of in-video reflective prompts, by asking them to compare learning experiences with and without prompts.

Study Design & Procedure

Participants watched two 8 minute lecture videos, one with reflective prompts, and one without any prompts. The videos were from Khan Academy, titled "Population standard deviation" and "Logarithms", respectively. The order of presentation and pairing of prompting condition with topic were counterbalanced.

After participants watched both videos, they were asked open-ended questions about their experience in either condition, to compare their experience with respect to enjoyment, cognitive goal, and the perceived benefits to learning. [1] The prompting condition was counterbalanced over all four prompting strategies (see Table 1a), which we discuss in more depth in Study 2.

Participants

We recruited 100 participants (55 male, mean age 35.1) on Amazon Mechanical Turk, paying $6 for an hour-long study. Recruiting crowd workers enables us to (1) obtain a more general population compared to lab settings and (2) have greater experimental control to collect more extensive data about learning experiences, even though the motivation and background knowledge of crowd workers may be different from those of learners in online learning platforms.

[1]We also administered pretests and posttests to measure learning, but did not see significant effects.

Table 2: Main pros and cons of in-video prompting from learners' perspective (number of mentions in parenthesis)

	Pros	Cons
Concentration	- Enhance learners' concentration. (42)	- Distract from the learning process. (59)
Learning process	- Encourage reflection. (57) - Split the lecture into small pieces. (16) - Help grasp key concepts. (10)	- Provide no feedback on responses. (9)
Emotional responses	- Provide interactivity. (5)	- Cause anxiety. (17)

Qualitative Results

To analyze participants' comments, we first separated comments consisting of more than one point into multiple comments so that each comment contains only one idea. We then categorized each comment into two groups based on whether it was positive or negative. Out of 231 comments, 130 were positive and 101 were negative. For comments in each group, one researcher conducted open coding to categorize each response. Afterwards, another researcher verified the coded labels and resolved conflicts with discussion. Table 2 describes the reported pros and cons of in-video prompting from learners' perspective. The extracted categories of participants' comments are as follows:

Enhance learners' concentration. Prompting encourages learners to pay more attention to the video. Responding to the question ensures that learners are on the right track, which makes learners more engaged. A participant mentioned, *"the prompts gave me a good attention check to make sure I understood what was being discussed."*

Encourage reflection. Prompting asks learners to reflect what they have learned. It reinforces the knowledge, which leads to a better learning experience. A participant wrote, *"it helps you think about what you know and have learned through the video, making you actively recall and cement your learning right away."*

Split the lecture into small pieces. Prompting splits the knowledge to digest in a more fine-grained way. By forcing learners to stop and think at checkpoints, the amount of knowledge to absorb at once is reduced: *"With prompts, the learning is broken down into stages and you have to think and reiterate what you've learned so far, which makes it easier to remember."*

Help grasp key concepts. Prompting helps learners to grasp the most important concepts of the lecture. Participants perceived the moment of prompting as important checkpoints, which makes learners think about the most important concepts they have learned so far. A participant pointed out, *"Having a prompt helps to indicate exactly what the key concept was so you can take a moment and decide if you fully understand it."*

Provide interactivity. Prompting is an interactive activity. Participants said they enjoyed the interactivity and felt they have learned more. This sentiment is echoed by a participant: *"I think prompts make videos more hands-on and interactive and deliver a more educational experience."*

Distract from the learning process. Prompting might break the flow of concentration. Forcing learners to respond to the questions even if they are following the lecture very well can lead to a negative learning experience. A participant noted, *"it might cause you to lose focus on the material in the video by breaking your chain of thought because you are basically being interrupted."*

Provide no feedback on responses. Learners wish to receive feedback on their response to make sure that they properly respond to the questions. Without feedback, learners may be less motivated to respond to prompts. A participant commented, *"there is no feedback so even if I answer the prompt question and I'm confident, I may be wrong."*

Cause anxiety. Some learners feel worried about giving inappropriate or inadequate responses. They feel the responses are being monitored, which makes learners frustrated when they struggle to come up with an appropriate response: *"I felt frustrated that it appeared difficult for me to explain what I learned thus far."*

STUDY 2.   EFFECTS OF DIFFERENT PROMPTING STRATEGIES

Study 2 investigated how different kinds of prompts might be perceived by learners, and collected quantitative measures of how learners perceived prompts.

We investigated two dimensions of prompting questions: the comprehension-experience orientation and the level of specificity. The comprehension-experience orientation represents which information the question seeks to reveal. Comprehension-centered questions ("Describe what you have learned so far.") asked learners to reflect on the contents of the lecture and their current comprehension. These questions promote self-explanation in learners, which previous research suggests could be beneficial for their learning [24, 5]. Responses to comprehension-centered questions allow instructors to identify how well learners are following the lecture. Experience-centered questions ("Describe something unsatisfying about the lecture so far.") reveal learning experiences during the lecture video. By directly asking what makes learners unsatisfied, instructors can collect actionable feedback on their instruction.

The level of specificity determines whether the prompting question refers directly to lecture content. General questions ("Describe something unsatisfying about the lecture so far.") do not refer to the content of the lecture. These questions

could be shown anywhere in a given video. Specific questions ("Describe how to calculate a mean.") refer to concepts in the lecture video. Instructors should consider what to ask at specific moments, which requires more effort for instructors to build the prompting questions. Table 1a shows the design space of prompting questions that we cover, and examples of each.

In addition, we investigated whether perceptions of prompt types might vary based on learners' prior knowledge of the video content.

We investigated the following research questions.

- RQ2a. How does the effect of in-video prompting vary according to the kind of prompting strategy?

- RQ2b. How do learners with different levels of achievement perceive each prompting strategy?

Study Design
The study used a between-subjects design, where each student was randomly assigned to one of the four prompting strategies described in Table 1a. Each participant watched a lecture video from Khan Academy on "Population standard deviation". We gave prompts at the beginning, middle, and end of the video. We denote each condition using four-letter codes; Co-Ge for the comprehension-centered and general condition, Co-Sp for comprehension and specific, Ex-Ge for experience and general, and Ex-Sp for experience and specific.

To understand the differences between prompting strategies, we included quantitative measures of learners' experiences. After watching the video, learners were asked to rate their agreement on a seven points scale with six statements about their experience. For example, a learner would be asked to rate on a scale from 1 (Strongly Disagree) to 7 (Strongly Agree) whether "Prompting helped me to pay attention to the lecture" (Q1). The six statements about experience with prompting are listed in Figure 2. These asked about the extent to which learners agreed or disagreed that prompting helped them pay attention to the video, understand, grasp the most important ideas, was enjoyable, interrupted their learning process, or made them worried about giving inappropriate responses. These statements were chosen by using the qualitative dimensions identified in Study 1.

Participants also answered two questions about their cognitive load while watching the video [12]. They rated on a sliding scale from 0 to 100: "How much mental demand did you experience watching this lecture?" and "How much effort did it take you to watch this lecture?".

Participants
We recruited 200 participants on Amazon Mechanical Turk, paying $3.5 for a 35-minute long study. Participants' mean age was 36.8 (SD = 11.5; min = 20; max = 72) with 102 male. For each subgroup by prompting strategy (Co-Sp, Co-Ge, Ex-Sp, Ex-Ge), the mean ages were 39.2, 36.1, 37.3, and 36.6, respectively.

Study Procedure
Participants were asked to (1) take a pretest with six problems, (2) watch a video with prompting, (3) respond to a survey related to the video watching experience, and then (4) take a posttest. The session ended with a final survey asking for participants' general experiences in free-form text.

Results
Different prompting strategies had different effects on participants' perceived learning experience. As shown in Figure 2, learners rated comprehension-centered prompts as more helpful than experience-centered prompts on average, as measured by judgments on the positive questions Q1-Q4 (2-way ANOVA, $F(1, 196) = 10.04$ and $p < 0.005$, $F(1, 196) = 33.11$ and $p < 0.0001$, $F(1, 196) = 33.00$ and $p < 0.0001$, and $F(1, 196) = 16.67$ and $p < 0.0001$ for Q1, Q2, Q3, and Q4 respectively). Participants in the comprehension-centered conditions mentioned that they could take a breather, have time to reflect on what has been learned, and assess their understanding on their own. One participant remarked, *"It was a way to help ensure I understood the whole process by breaking it down with questions to answer, instead of trying to absorb it all at once and remember everything after."*

Participants rated the experience-centered prompts as more distracting (2-way ANOVA, $F(1,196)=17.13$ and $p<0.0001$ for Q5). Participants in the experience-centered conditions mentioned that it was hard to learn from the video while at the same time trying to give comments on how the instructor could improve the video. One participant remarked, *"Prompting had me thinking about many things at once. This caused me to lose focus."* However, there is also a bright side of experience-centered prompts, especially regarding the learners' emotional experience. Participants said that they liked being able to leave a subjective comment to the instructor and it made them feel involved in the lecture.

Although participants could take advantage of comprehension-centered prompts in their learning, being asked about the contents of the lecture in depth irritated some learners. On the other hand, experience-centered prompts ask about participants' subjective opinion regarding the lecture and therefore it can be thought to be less pressure for learners. Our survey results show that, however, there is no such difference between comprehension-centered and experience-centered prompting strategies (Q6).

The effect of learners' prior knowledge
The qualitative comments suggest that learners differ in their opinions about the value of prompting. To understand the effect of each prompting strategy for learners with unequal prior knowledge, we divided each group into two subgroups based on participants' pretest scores and discovered how the experience differed for each subgroup. As the cut-off point to separate the two groups we used a score of 2 out of 6, the median pretest score of all participants. Table 3 shows the number of participants in each subgroup.

Figure 3a illustrates self-reported cognitive load of each group for four different prompting conditions. The high-score group perceived significantly less cognitive load under the general
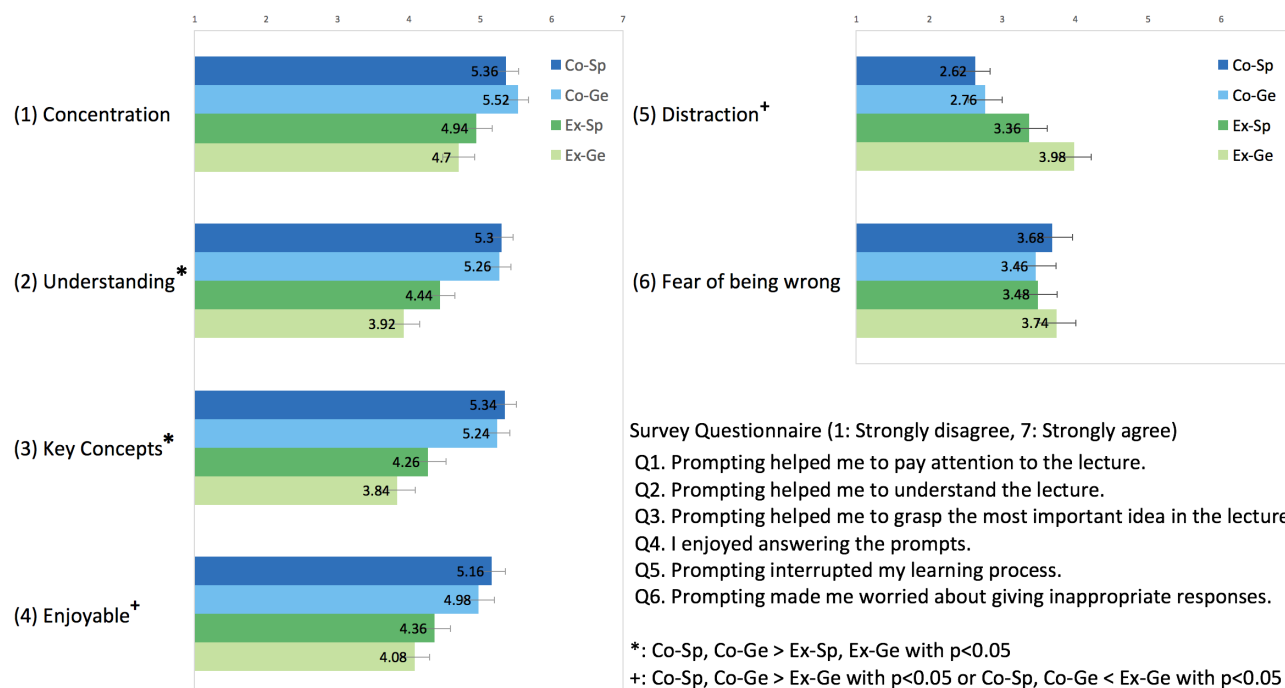
Figure 2: Participants' self-reported learning experiences captured by responses to 6 survey questions (Q1-Q6). Bar graphs show the responses to each question for each of the four prompting strategies. Questions on the left column (Q1-Q4) address positive experiences while questions on the right column (Q5-Q6) address negative experiences. Error bar: +/- 1 standard error of the mean.

Table 3: The number of participants in each group and subgroup. Co-Sp: comprehension-centered and specific, Co-Ge: comprehension-centered and general, Ex-Sp: Experience-centered and specific, and Ex-Ge: experience-centered and general. High: participants with high pretest scores; Low: participants with low pretest scores.

	Total	High	Low
Co-Sp	50	22	28
Co-Ge	50	21	29
Ex-Sp	50	23	27
Ex-Ge	50	31	19

prompting conditions (ANOVA, $p<0.05$ for both conditions). This result corresponds to the findings from earlier research [2] that high-performing students are good at providing an answer to generic questions.

Figure 3b shows how each subgroup perceived anxiety for different prompting conditions. The result indicates that the low-score group felt more anxiety than the high-score group under specific prompting condition and the opposite pattern is observed for the general prompting conditions. This outcome may seem incongruous with the previous finding that the high-score group exhibit lower cognitive load with general prompts.

However, study data collected from this experiment could provide alternative explanations for this result. First, a number of participants in the low-score group provided simple responses under general prompting conditions (e.g., "population standard deviation" for the prompt "Describe what you have learned so far"), leaving less chance of being wrong. In addition, with the experience-general prompts, many high-score participants mentioned it was hard for them to find unsatisfying points in the lecture, increasing their anxiety regarding their responses.

The result indicates that the high-score group and the low-score group perceived a different level of cognitive load and anxiety. The low-score group reported higher cognitive load than the high-score group and the differences were especially large for general prompting conditions. Regarding the level of anxiety, the low and high-score group behaved differently for the specific and general conditions.

LEARNER RESPONSES AS FEEDBACK TO INSTRUCTORS

We conducted a series of interviews with instructors and instructional designers to understand the usefulness of learner responses as feedback. Instructors and instructional designers are important stakeholders in in-video prompting, as they are the ones who author the prompts and potentially benefit from the collected learner responses. This section presents results from the interview study with instructors.
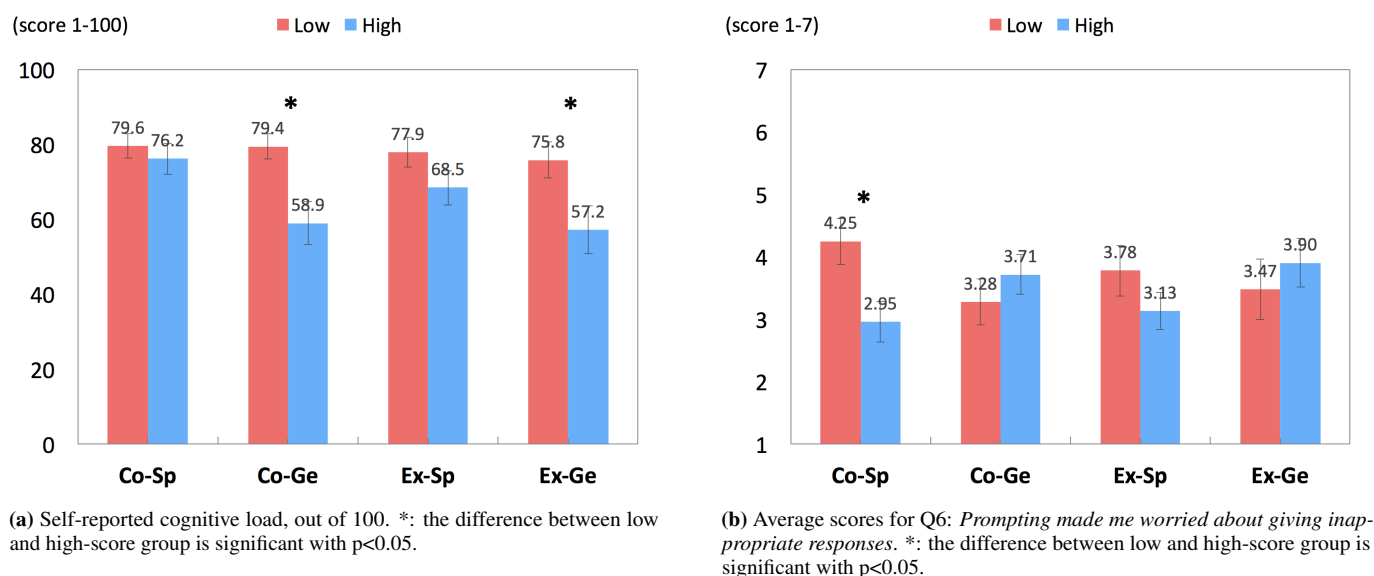
(a) Self-reported cognitive load, out of 100. *: the difference between low and high-score group is significant with p<0.05.

(b) Average scores for Q6: *Prompting made me worried about giving inappropriate responses*. *: the difference between low and high-score group is significant with p<0.05.

Figure 3: Perceived experience for Low and High group in each prompting condition. (a) The low-score group reported higher cognitive load than the high-score group when they encountered general prompts. (b) The low-score group reported higher anxiety than the high-score group when they were given a comprehension-centered and specific prompt. Error bar:+/-1 standard error of the mean.

Procedure

We had 1-hour long interviews with instructors asking how useful the collected responses to in-video prompts from learners might be as feedback. We recruited 3 instructors who had experience in publishing lecture videos. Two instructors had published 8 hours and 10 hours of lecture video in total, respectively. The other instructor had led a flipped classroom for 8 semesters. After explaining the prompting strategies, we presented the learner responses as well as the corresponding questions collected in Study 2 and asked how they would make use of the responses as feedback. Instructors explored the learner responses for 10-15 minutes. After the exploration, we asked questions about using the responses as feedback.

Results

Specificity of learners' responses. Instructors found that the collected responses are helpful as feedback because the responses were specific. They noted that asking questions in the middle of a video yields more specific responses. An instructor remarked, *"If instructors ask for comments at the end of the lecture, it is hard for learners to give comments on specific points of the lecture because they are likely to have forgotten details."*

Needs for organizing learners' comments. Instructors expressed needs for clustering the responses, based on criteria such as whether the comments are about visuals, pronunciation, or lack of information. An instructor said, *"I thought it would be easier to read the comments if they were organized."*

Comprehension-centered vs. Experience-centered. Instructors responded that experience-centered questions yield more actionable feedback than comprehension-centered questions. The responses from experience-centered questions (e.g., *"The instructor's handwriting is a little bad"*) tend to point out the problems on instructional delivery, but the responses from comprehension-centered questions (e.g., *"The standard deviation is calculated by taking the square root of the variance."*) describe their comprehension. Instructors were not sure whether the responses from comprehension-centered questions reflect learners' true level of comprehension. An instructor noted, *"If learners' responses are not good in comprehension-centered questions, it is hard to know whether the learner doesn't know, or the learner is just tired."*

Instructors had different opinions on which questions help learning more. Two instructors said asking comprehension-centered questions is more helpful for learners because they promote reflection on the lecture while describing learning experience is not quite relevant to learning. One of the instructors said, *"I think learners feel like unnecessarily responding to experience-centered questions, and feel like checking their understanding when they respond to comprehension-centered questions."* However, another instructor mentioned experience-centered questions help learning more: *"I think experience-centered responses are going to inherently be more specific than comprehension-centered. [. . .]When we are talking about learning a particular topic, it's just going to be way more useful to talk about the thing that we can both see clearly between the two of us [. . .]. That's going to be a more productive discussion than what's going on inside my head."*

Specific vs. General. Instructors observed that different levels of specificity in prompts result in different types of feedback. For experience-centered questions, instructors said responses to general questions could serve as feedback on instruction delivery (e.g., *"It was good, just a little slow"*), such as the speed of speech, but responses to specific questions

could serve as content-related feedback (e.g., *"if the entire equation was done in meters, I would have a better and easier understanding of how it works"*), such as the clarity of a particular explanation.

INSTRUCTIONAL DESIGNERS' VIEW ON RESPONSES
To understand the usefulness of the responses as feedback from an educational point of view, we interviewed instructional designers.

Procedure
Similar to the interview study with instructors, we presented the responses as well as corresponding questions. Then we explained the prompting strategies and the instructional designers explored the responses. We conducted 1-hour long interviews with three instructional designers who are managing online courses on campus, including those on Coursera.

Results
Instructional designers commented that the responses could serve as feedback to instructors, identified problems in our current prompts, and suggested ways to improve the design of prompting strategies.

Feedback to instructors. Instructional designers said that the responses are useful as self-checklists for instructors, but to improve the lecture, the responses should be coupled with instructional design components that the instructor should consider in the lecture. Also, they remarked that high-level feedback such as comments on learning objectives and organization of presentations is more useful for instructors.

Importance of specific questions. Upon inspecting the current question prompts, instructional designers identified issues in the questions. They mentioned some questions were not concrete enough and too superficial to diagnose issues with the lecture. Designers suggested including more specific questions such as "Is the pitch of the instructor's voice appropriate?".

Mixing multiple prompting strategies. Instructional designers suggested using multiple prompting strategies even in a single video. They said asking experience-centered questions multiple times during a single lecture video could distract learners and damage the learning experience. Presenting experience-centered questions at the end and comprehension-centered questions in the middle of the video could be effective for both learners and instructors.

DISCUSSION
In this research, we attempt to understand the effect of in-video prompting, an under-explored yet complex topic. In-video prompting has potential to make video-based learning more interactive and even increase learning, while providing instructors with valuable data. Our goal is not to show that a particular prompting strategy is better than others or to conclude that prompting should be designed in a particular way for all videos, but rather (1) to contribute an in-depth understanding of differing perspectives on in-video prompting between learners, instructors, and instructional designers, and (2) to explore the design space of in-video prompting and the trade-offs involved. We discuss several issues in in-video

prompting that involve handling trade-offs and making design decisions.

Trade-off Between Learners and Instructors
One observation from our studies was that learners and instructors might prefer different prompting strategies for different reason. Learners perceive comprehension-centered prompting as enjoyable, less interrupting, and helping them learn. However, instructors generally found that the responses from experience-centered questions to be more actionable, which makes it easy for instructors to figure out the points that need to be improved. With this trade-off in mind, prompting strategies should be carefully designed to maximize the benefits for both stakeholders.

One way to address the trade-off is to design a hybrid prompting strategy, in which multiple prompt types are presented to a learner while watching a video, as suggested by instructional designers we interviewed. However, several issues could arise. It might be the case that drawbacks of both prompt types could be observed. Moreover, presenting both prompt types makes learners respond to two different types of questions in a lecture video, which may increase their cognitive load. Further study is needed to determine whether and how this hybrid prompting strategy helps learners.

Trade-off Between Questions and Responses for Instructors
Designing specific prompts is more expensive than designing general prompts because specific prompts needs to be coupled with lecture content. For the responses, however, specific prompting tends to yield more specific responses, which instructors might find useful. Instructors should choose the level of specificity of the questions, but it is not a straightforward task. Also, specific questions could yield responses that are too narrow, which could compromise the diversity of responses.

It is important to balance the level of specificity to get diverse but insightful responses. One way to meet the balance is to design multiple specific questions. For example, we currently consider content-related specificity, but we also can consider instructional delivery-related specificity, such as the pace of lecture, the tone of voice, and the speed of speech. By considering multiple types of specificity in parallel in their prompt design, instructors could potentially receive diverse and specific responses.

Learner Feedback from Multiple Points of View
Through the interviews, we observed that the meaning of useful feedback is different between instructors and instructional designers. Instructors found it useful to get actionable feedback, whereas instructional designers found it useful to get more high-level feedback.

This leads us to think about what good feedback is for instructors and instructional designers. We demonstrated how collecting actionable comments from learners is possible, but making sense of them, finding patterns in them, and deriving high-level points require extra work. Future work could investigate collecting and processing the responses of learners to generate more high-level, aggregate feedback to instructors.

Exploring the Design Space of In-video Prompting
It is hard to say that one prompting strategy always outperforms another. As discussed above, there is a trade-off along the dimensions of prompting strategies for different stakeholders. However, there still remains much room for improvement in designing effective prompts.

The effects of in-video prompting highly depend on (1) who the learners are, (2) how difficult or well-structured the video is, and (3) what the prompts ask about. For example, as we observed in our study with crowd workers, prompting for negative feedback will generate only meager responses if the video is already well-structured, while learners may have a hard time writing their response. Likewise, requiring too detailed knowledge in in-video prompting may frustrate low-performing learners, who are already prone to drop out. In an online learning environment, where thousands of videos meet millions of learners, thoughtfully designed and adjusted in-video prompting has potential to provide significant benefits to both learners and instructors. Future work could explore the feasibility of dynamically adjusting the prompting strategy for each video (content-specific), or generating a personalized prompt plan based on learners' course interaction (learner-specific).

LIMITATIONS AND FUTURE WORK
This work aims to provide guidance to instructors and researchers about how learners perceive in-video prompts, and when it might be worth including them, and in what form. Due to the exploratory nature of this work, however, it leaves several points unanswered.

We did not investigate how in-video prompting affects learning outcomes since our goal is not to yield a clear conclusion about how effective in-video prompting is. As research shows that reflective prompting could yield learning gains [5, 24], future work could address the relationship between prompting strategies and learning outcomes.

This paper only covers a subset of dimensions in designing prompting strategies: comprehension-experience orientation and the level of specificity. For the prompt positions in our studies, the rationale followed the affordance of time-anchored video prompting; the positioning of reflective prompts afford different levels of specificity, e.g., a prompt at the end of a video could ask to reflect more generally, whereas a prompt in the middle could refer to the specific concept just covered. We recognize that there might be other important design factors not addressed in this work. Future work could address other dimensions such as frequency of prompting and mode of learner responses.

We collected responses from crowd workers, which could bias results, e.g., due to selection bias or non-representativeness. It is not clear whether our findings might generalize to learners in online learning platforms. Moreover, the crowd workers responded to all the prompting questions partly because of monetary reward. Future work should perform a deployment study to test the effect of in-video prompting in real-world educational settings.

Designing personalized prompting strategies could be an interesting future work. By leveraging learners' data, we could choose the most beneficial prompting strategies. For example, if the learner is already good at the subject covered by the lecture, the instructor could provide prompting with more challenging questions.

Future work could also investigate the design of the instructor dashboard. Interview results suggest that instructors have needs for efficiently organizing and exploring the collected learner responses to easily grasp the overall perception of comments. Automatically organizing, aggregating, and visualizing learner responses could be an interesting future direction to explore.

CONCLUSION
This paper investigated the effects of in-video prompting on both learners and instructors. In-video prompting enables instructors to collect specific comments from learners. To understand how in-video prompting affects the learning experience, we conducted two studies with crowd workers. Results showed that different prompting strategies have different effects on the learning experience. Learners perceive that learning-centered questions are less interrupting, more enjoyable, and more helpful for learning. Interviews with instructors revealed that in-video prompting gives specific comments to them and that responses from experience-centered questions are more actionable. Instructional designers emphasized the importance of coupling question design with instructional design components for more useful feedback.

REFERENCES
 1. Akshay Agrawal, Jagadish Venkatraman, Shane Leonard, and Andreas Paepcke. 2015. YouEDU: addressing confusion in MOOC discussion forums by recommending instructional video clips. (2015).

 2. Vincent Aleven, Niels Pinkwart, Kevin Ashley, and Collin Lynch. 2006. Supporting self-explanation of argument transcripts: Specific v. generic prompts. In *Workshop of Intelligent Tutoring Systems for Ill-Defined domains, 8th International Conference on Intelligent Tutoring Systems*. 47–55.

 3. Lori Breslow, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and Daniel T Seaton. 2013. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment* 8 (2013).

 4. Yuanzhe Chen, Qing Chen, Mingqian Zhao, Sebastien Boyer, Kalyan Veeramachaneni, and Huamin Qu. 2016. DropoutSeer: Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and

prediction. In *Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on*. IEEE, 111–120. `DOI: http://dx.doi.org/10.1109/VAST.2016.7883517`

5. Michelene TH Chi, Nicholas Leeuw, Mei-Hung Chiu, and Christian LaVancher. 1994. Eliciting self-explanations improves understanding. *Cognitive science* 18, 3 (1994), 439–477. `DOI: http://dx.doi.org/10.1207/s15516709cog1803_3`

6. Clark A Chinn and William F Brewer. 1993. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of educational research* 63, 1 (1993), 1–49. `DOI: http://dx.doi.org/10.3102/00346543063001001`

7. John F Chizmar and Anthony L Ostrosky. 1998. The one-minute paper: Some empirical findings. *The Journal of Economic Education* 29, 1 (1998), 3–10. `DOI: http://dx.doi.org/10.1080/00220489809596436`

8. K Patricia Cross and Thomas A Angelo. 1988. Classroom Assessment Techniques. A Handbook for Faculty. (1988).

9. Elena L Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. Mudslide: A spatially anchored census of student confusion for online lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1555–1564. `DOI: http://dx.doi.org/10.1145/2702123.2702304`

10. Jeffrey D Karpicke and Janell R Blunt. 2011. Retrieval practice produces more learning than elaborative studying with concept mapping. *Science* 331, 6018 (2011), 772–775. `DOI: http://dx.doi.org/10.1126/science.1199327`

11. Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen Daniel Li, Krzysztof Z Gajos, and Robert C Miller. 2014. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 563–572. `DOI: http://dx.doi.org/10.1145/2642918.2647389`

12. René F Kizilcec, Jeremy N Bailenson, and Charles J Gomez. 2015. The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107, 3 (2015), 724. `DOI: http://dx.doi.org/10.1037/edu0000013`

13. Geza Kovacs. 2016. Effects of in-video Quizzes on MOOC lecture viewing. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 31–40. `DOI:http://dx.doi.org/10.1145/2876034.2876041`

14. Yi-Chieh Lee, Wen-Chieh Lin, Fu-Yin Cherng, Hao-Chuan Wang, Ching-Ying Sung, and Jung-Tai King. 2015. Using time-anchored peer comments to enhance social interaction in online educational videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 689–698. `DOI:http://dx.doi.org/10.1145/2702123.2702349`

15. Margaret Mazzolini and Sarah Maddison. 2007. When to jump in: The role of the instructor in online discussion forums. *Computers & Education* 49, 2 (2007), 193–213. `DOI:http://dx.doi.org/10.1016/j.compedu.2005.06.011`

16. Samuel T Moulton, Selen Türkay, and Stephen M Kosslyn. 2017. Does a presentation's medium affect its message? PowerPoint, Prezi, and oral presentations. *PloS one* 12, 7 (2017), e0178774. `DOI: http://dx.doi.org/10.1371/journal.pone.0186673`

17. Amy Pavel, Dan B Goldman, Björn Hartmann, and Maneesh Agrawala. 2016. VidCrit: Video-based Asynchronous Video Review. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 517–528. `DOI: http://dx.doi.org/10.1145/2984511.2984552`

18. Daniel L Schwartz and Taylor Martin. 2004. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction* 22, 2 (2004), 129–184. `DOI: http://dx.doi.org/10.1207/s1532690xci2202_1`

19. Vikash Singh, Sarah Abdellahi, Mary Lou Maher, and Celine Latulipe. 2016. The Video Collaboratory as a Learning Environment. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. ACM, 352–357. `DOI: http://dx.doi.org/10.1145/2839509.2844588`

20. Kristin Stephens-Martinez, Marti A Hearst, and Armando Fox. 2014. Monitoring moocs: which information sources do instructors value?. In *Proceedings of the first ACM conference on Learning@ scale conference*. ACM, 79–88. `DOI:http://dx.doi.org/10.1145/2556325.2566246`

21. Richard L Weaver and Howard W Cotrell. 1985. Mental aerobics: The half-sheet response. *Innovative Higher Education* 10, 1 (1985), 23–31. `DOI: http://dx.doi.org/10.1007/BF00893466`

22. Miaomiao Wen, Diyi Yang, and Carolyn Rose. 2014. Sentiment Analysis in MOOC Discussion Forums: What does it tell us?. In *Educational data mining 2014*.

23. Joseph J Williams and Tania Lombrozo. 2010. The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science* 34, 5 (2010), 776–806. `DOI: http://dx.doi.org/10.1111/j.1551-6709.2010.01113.x`

24. Joseph Jay Williams, Tania Lombrozo, Anne Hsu, Bernd Huber, and Juho Kim. 2016. Revising Learner Misconceptions Without Feedback: Prompting for Reflection on Anomalies. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 470–474. `DOI: http://dx.doi.org/10.1145/2858036.2858361`

25. Joseph Jay Williams, Tania Lombrozo, and Bob Rehder. 2013. The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General* 142, 4 (2013), 1006. `DOI: http://dx.doi.org/10.1037/a0030996`

26. Han Zhang, Maosong Sun, Xiaochen Wang, Zhengyang Song, Jie Tang, and Jimeng Sun. 2017. Smart Jump: Automated Navigation Suggestion for Videos in MOOCs. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 331–339. DOI: `http://dx.doi.org/10.1145/3041021.3054166`

27. Saijing Zheng, Pamela Wisniewski, Mary Beth Rosson, and John M Carroll. 2016. Ask the Instructors: Motivations and Challenges of Teaching Massive Open Online Courses. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 206–221. DOI: `http://dx.doi.org/10.1145/2818048.2820082`