EduLive: Re-Creating Cues for Instructor-Learners Interaction in Educational Live Streams with Learners' Transcript-Based Annotations

JINGCHAO FANG, University of California, Davis, USA
JEONGEON PARK*, DGIST, Republic of Korea
JUHO KIM, KAIST, Republic of Korea
HAO-CHUAN WANG, University of California, Davis, USA

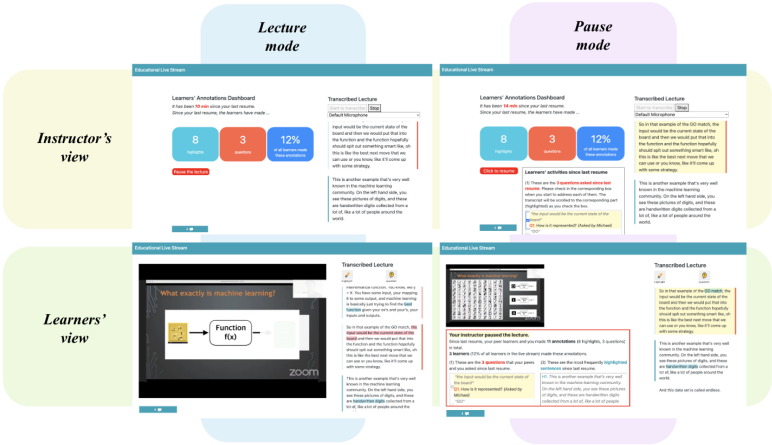

Fig. 1. An overview of EduLive's two views and two modes. Both views are switched from the lecture mode to the pause mode when the instructor decides to pause to teach new material and clicks the "pause" button. In the pause mode, a summary box is shown on both views which details learners' transcript-based annotations.

Educational live streaming has become a complement to in-person teaching. While synchronous instructor-learner communication is useful, the technology-mediated nature of live streaming can obscure many interaction cues (e.g., learners' facial expressions and body language), which dampens the instructors' ability to respond to remote learners' needs. We explore the opportunity of leveraging real-time transcripts generated from instructors' audio as a basis for re-creating interaction cues. Transcripts can be leveraged to reveal the content of live streams in a form that learners can trace back and annotate, and such annotations can be further aggregated and presented to instructors as signals to assist them in tracking learners' engagement. By designing and evaluating our proof-of-concept prototype system, EduLive, we show that instructors

*Work done while at KAIST.

Authors' Contact Information: Jingchao Fang, jcfang@ucdavis.edu, University of California, Davis, Davis, California, USA; Jeongeon Park, jeongeonpark1@gmail.com, DGIST, Daegu, Republic of Korea; Juho Kim, juhokim@kaist.ac.kr, KAIST, Daejeon, Republic of Korea; Hao-Chuan Wang, hciwang@ucdavis.edu, University of California, Davis, Davis, California, USA.

ACM 2573-0142/2024/11-ART421
https://doi.org/10.1145/3686960

benefited from the summative information extracted from learners' annotations, and the context provided by the transcript enhanced their ability to answer learners' questions. Our system contributes to the design space of social annotations in CSCW by employing social annotations in educational live streaming scenarios.

CCS Concepts: • **Human-centered computing → Interactive systems and tools**.

Additional Key Words and Phrases: Educational live streaming, transcript-based annotation, instructor-learners interaction, educational technology, online learning, social annotation, interactive interface

ACM Reference Format:
Jingchao Fang, Jeongeon Park, Juho Kim, and Hao-Chuan Wang. 2024. EduLive: Re-Creating Cues for Instructor-Learners Interaction in Educational Live Streams with Learners' Transcript-Based Annotations. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 421 (November 2024), 33 pages. https://doi.org/10.1145/3686960

1 Introduction

Educational live streams, in which instructors and learners engage in the knowledge transfer process synchronously, provide accessible and scalable learning opportunities to many people. While asynchronous teaching or learning by posting and watching pre-recorded videos may cause difficulty in interacting with each other [31], educational live streams conducted through streaming platforms and videoconferencing software could, theoretically, support both instructor-student interaction and peer interaction in real-time. For example, Zoom [1] allows every participant to turn on the microphone and camera to speak, or enter into breakout rooms for peer interactions in small groups. YouTube Live [2] and Twitch [3] enable audiences to communicate with the streamer by sending text messages and emojis in the chatroom.

Yet, recent studies revealed that educational live streaming is not as effective in mitigating the reduced interactivity of online learning as people expected. Yarmand et al. showed that learners were reluctant to turn on their cameras when attending classes through Zoom, and struggled to connect with instructors and peers due to the lack of post-lecture time for informal exchanges, which led to a lack of sense of community and feeling isolated [67]. Text-based interactions through chatrooms are also not ideal since questions or comments can easily get ignored in the flow of the text messages [25, 65]. Instructors of educational live streams also face a number of challenges, including the difficulty of sensing learners' engagement and confusion when commonly used cues in instructor-learner interaction such as eye gaze and facial expression were limited online [10, 67]. Even though some of these cues might be made available with learners' videos, they tend to be too nuanced for the instructor to digest and offer no directly actionable advice for teaching, especially when the number of students online is excessive. Instructors also expressed the need for seeing aggregated cues and signals that can assist them in "reading the class" [67].

Crowdsourcing and aggregating learners-generated annotations could be a way to collect such cues to aid online teaching from the perspective of instructor-learners communication. We are inspired by the notion of social annotations, defined as "freely established associations between resources and metadata (keywords, categories, ratings) performed by a community of users with little or no central coordination" [6]. Employing social annotations such as highlights in articles and photo tags marked by former viewers has become a commonly adopted approach in social computing systems for aggregating social inputs to support information consumption (e.g., [38]) and knowledge sharing (e.g., [35]). The practice of crowdsourcing learners' generated annotations could be conceptualized as *active learnersourcing*, in which learners make small contributions through a set of pedagogically beneficial activities while the system collects useful information

[1] https://zoom.us/
[2] https://www.youtube.com/
[3] https://www.twitch.tv/

as a byproduct [33]. Previous research showed that learners-generated annotations can enhance peer learners' presence in learning communities [18]. In the scenario of educational live streaming, while interaction cues available in face-to-face communication (e.g., eye gaze and body language) can be obscured by the absence of audio and visual representations, learners-generated annotations, after being processed and aggregated, may provide instructors with cues and signals about learners' presence and engagement, helping them to read the class and suggest actions to address learners' needs and feedback. In addition, annotating learning material during educational live streams (e.g., highlighting key concepts or marking confusion) is a meaningful practice for learners that scaffolds their cognitive processing of the content.

However, annotating live streaming content is challenging. One key obstacle arises from the limited reviewability inherent in live streaming videos. Typically, learners wish to annotate content based on previously presented material, as it takes time to process the information before they can formulate a concrete question. Yet, some live streaming tools (e.g., Zoom) lack support for video rewind. While other tools (e.g., YouTube Live) permit audience rewinding during live streams, reviewing video segments and making annotations on top of them inevitably leads to missing information that is delivered by the streamer at the current moment. Similarly, video-based annotations are hard for instructors to handle, as they have to review the video segment to understand the context of learners' annotations. The *same modality* of the currently presented content and the content to be reviewed and annotated caused a conflict. In other words, audiences cannot watch two videos at the same time. One way of mitigating this conflict is to transform the antecedent video segments into a different modality, so that the audiences can reach the previously streamed content for reviewing and annotating, without completely losing access to the content being delivered at the moment.

We propose to provide real-time transcripts of live streams as the basis for learners to annotate. First of all, plain text is one of the most accessible modalities for people to annotate on in existing social computing applications (e.g., text-based highlights on Medium articles [4] and Coursera video transcripts [5]) and thus we expect most online learners to find it straightforward to engage with the practice. Second, transcripts can capture the major content of the live streams since much information in educational live streams is conveyed through speech [50]. Third, transcript-based annotation is a relatively lightweight task that is not expected to distract learners much from learning [19]. Most importantly, transcripts can provide reviewability, allowing both instructors and learners to re-visit streamed content in order to generate or check annotations.

We introduce EduLive, a proof-of-concept prototype system for educational live streaming support. EduLive collects learners-generated transcript-based annotations (i.e., highlights and questions)and aggregates and presents these annotations in one of the two different granularities, depending on the current mode of the live stream. A pausing mechanism is introduced into EduLive to identify the mode of the live stream. Based on the amount of learners' annotations, EduLive suggests moments for instructors to "pause" their presentation to read and address learners' annotations, splitting the live streams into the "lecture mode" and the "pause mode". By switching between these two modes, instructors can dynamically shift between (a) concentrating on delivering new content while being provided with summative information for sensing the overall class engagement; realized by EduLive's low granularity presentation of learners' collective annotations, and (b) taking time to interact with learners by reacting to specific learners-generated annotations, made possible by EduLive's high granularity presentation. By asking learners to perform simple and pedagogically meaningful annotating tasks, EduLive acquires information to be reorganized

[4]https://medium.com/
[5]https://www.coursera.org/

into diverse formats (dashboard, annotations summary box, etc.), informing instructors of learners' collective presence and engagement and reshaping the interaction dynamics of the live streams. Figure 2 shows how annotations flow between learners and the instructor in EduLive. While we aim to broadly address diverse types of educational live streams, we expect that the design of EduLive is particularly relevant in tackling interaction challenges in webinar-like sessions hosted on streaming platforms (e.g., YouTube or Twitch), in which streamers primarily rely on audiences' text messages to sense their engagement and seek opportunities for instructor-learner interaction.

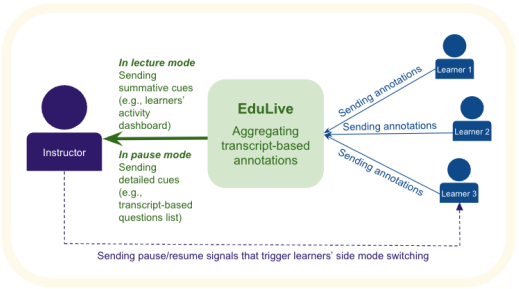

Fig. 2. The flow of transcript-based annotations and pause/resume signals in EduLive

By designing and deploying EduLive, we seek to answer the following research questions:

RQ1. Can learners' transcript-based annotations be aggregated to supplement instructors' awareness of learners' status and re-create interaction cues?

RQ2. How do instructors and learners make use of transcripts and aggregated annotations in EduLive to augment their teaching and learning in educational live streams?

We arranged 4 live streaming classes with 4 instructors and 12 groups of learners (112 in total). In each class, one instructor hosted three educational live streaming sessions with the support of EduLive and two baseline systems (namely, EduChat and EduScript which support basic chatroom functions and/or real-time transcript) to teach three groups of learners. Through semi-structured interviews, we found that instructors are generally positive about their teaching experience with EduLive. While the specific advantages perceived by the instructors tend to differ based on their teaching styles and preferences, instructors in all classes perceived that the switch between lecture mode and pause mode reduced the distraction caused by learners' detailed annotations. Instructors commented that the aggregated learners-generated annotations in EduLive supplemented their awareness of learners' presence and engagement. They also pointed out that questions asked based on transcripts provide the context of the questions so that they could better understand what the learners were confused about. Most learners were satisfied with the overall learning experiences augmented by EduLive as they actively engaged with the classes by taking advantage of the transcripts and making many annotations on top of them; although some of them expressed concerns about the inaccuracy and latency of the transcript, which could have dampened its effectiveness.

The contributions of this work are as follows:

- EduLive, a proof-of-concept prototype system that employs a real-time transcript and enables transcript-based annotations for re-creating interaction cues during an educational live stream.
- Results from four classes that reflect how instructors and learners leverage transcript and aggregated transcript-based annotations for educational purposes.

- Insights into future work around applying transcripts and transcript-based annotations in educational scenarios.

2 Background

2.1 Challenges of teaching and learning through educational live streams

One widely adopted form of synchronous online education is educational live streams, where online lectures are carried out through videoconferencing software (e.g., Zoom) or streaming platforms (e.g., YouTube and Twitch) [10, 57]. Educational live streams are unique in that the classes are more scalable, but the interactions are limited as most instructor-learner interactions take place in the text format. Previous work identified the lack of social cues and limited interaction channels as critical challenges of educational live streams [12, 63], which negatively affect both the instructors and the learners during the live streaming. For example, Chen et al. described missing interaction cues such as eye contact and body language as the biggest challenge for both parties [10]. Yarmand et al. specified that engagement and confusion are two types of cues that instructors use to understand the learners' needs, but are hard to perceive in live streams [67]. In addition, Ylirisku et al. stated that limited interaction cues weaken personal connections between instructors and learners [68].

In addition, due to limited communication channels in most live streams, the learners can hardly affect the pace of the online classes [11, 12]. When instructors focus too much on the live stream contents, they rarely answer or even realize learners' questions, resulting in cognitive gaps between instructors and learners and limited teaching efficiency of the instructors [12]. Furthermore, even if the instructors notice learners' questions or comments, they easily fail to understand the exact context of learners-generated content as chat messages cannot effectively capture it, especially when these questions or comments are referring to the instructor's speech delivered a while ago [11].

In summary, the instructors' teaching experiences and learners' learning experiences are mutually dependent on each other, and the interaction channel serves as the bridge in the education ecosystem. While the tie between the two ends of the knowledge transfer process seems to be weakened by the current form of computer-mediated online education, we highlight that it can, in contrast, be strengthened by the affordances provided by technological advancements such as speech-to-text services. In this work, we aimed to address the challenges in educational live streams by re-creating interaction cues using learners' transcript-based annotations and utilizing them to assist instructors in interacting with their learners.

2.2 Soliciting cues and feedback when presenting and teaching

In order to effectively interact with audiences, instructors and presenters need to sense and track audiences' status and engagement during their teaching or presenting. Existing work attempts to solicit cues and feedback from audiences either implicitly or explicitly to provide signals to instructors and presenters as teaching and presenting aids.

2.2.1 Implicitly sensing audiences' status. A common approach is to use sensors or recording devices to implicitly collect audiences' physiological data and behaviors to assist instructors or presenters before, during, and after their presentations. Systems such as Glancee [45], EngageMeter [27], and AttentivU [36] showed the potential of biofeedback in detecting learners' confusion, emotion, and engagement in real-time. In addition, previous research also attempted to apply computer vision techniques to video and audio recordings to track audiences' gaze [4, 13], body movement [3], gestures [8], facial expression [3, 58], and speech [3]. These sensor-based methods could collect learners' real-time status and fill some vacancies of direct interaction cues in educational settings without disturbing learners to ask for their explicit feedback.

2.2.2 Explicitly requesting learners to send active feedback. Another line of work explicitly collects audience-generated annotations as teaching and presenting aids to foster communication between instructors/presenters and learners/audiences . Zyto et al. showed a successful application of instructors adapting their teaching style to take students' comments into account on the reading material [69]. Mudslide aggregated and displayed learners' confusion with learner-generated muddy points [21]. Hamilton et al. [26] and Chen et al. [7] utilized multimodal interactions as interaction mediums to share context and enhance engagement between students and the instructor. A line of work also utilizes learner-generated comments

2.2.3 Utilizing class interaction data for awareness and instructional decisions. Previous work further investigates approaches to present learners' and audiences' interaction data to increase awareness of peers and instructors, and help instructors make better instructional decisions. PeerPresents proposed a student peer-feedback system where students can exchange more helpful feedback with less burden [54]. AffectiveSpotlight observed that analyzing and spotlighting expressive facial responses and head gestures of audiences in videoconferencing helped presenters increase awareness of the audience and led to more self-assessment of the quality of their talk to the audiences [48]. Xhakaj et al. conducted contextual inquiry interviews and found that instructors utilize class data to determine the best form of remedy. They also discussed different types of instructional decisions teachers can make with data, from class-level decisions to individual and group-level decisions, and provided suggestions for designing a teacher's dashboard [64]. Similarly, Holstein et al. conducted a case study with an intelligent tutoring system and found that instructors preferred presentations that helped them prioritize their time among students [29].

Compared to sensor-based methods in which learners passively send out biological data to be processed and presented to instructors, the explicit approach of collecting cues and feedback gives learners more freedom to decide what to or not to share with instructors. As learners perceive more agency when not being sensed by automated systems or algorithms, they are not likely to have heavy ethical concerns as they do not perceive pervasive surveillance. Furthermore, collecting annotations does not require additional sensors or devices, which are not accessible to many instructors and learners in real-world educational settings.

Following the second line of work, we explored the design opportunities of supporting instructor-learner interactions by enabling generating and viewing of real-time transcript-based annotations during educational live streams. In addition, we explored presentations of aggregated and reorganized real-time transcription-based annotations and their effects on learners and instructors. Our presentations are focused on showcasing the collective status and feedback of the class, instead of focusing on supporting a few students as our setting is on educational live streaming.

2.3 Learnersourcing annotations to generate meaningful content

Learnersourcing, a form of crowdsourcing that engages learners in collective content production that benefits future learners while participating in meaningful learning activities themselves [33, 61], is widely adopted in many existing education contexts. The main key to success in implementing learnersourcing in a task is to produce a meaningful artifact that integrates the learner's activity, while ensuring that the activity itself is pedagogically meaningful [56].

Existing work explored ways of learnersourcing annotations to generate meaningful content such as explanations [62], recommendations of learning activities [32], and content creation [28] to benefit others such as subsequent learners . Gordon et al. developed a crowdsourcing workflow using Codepourri, a system that creates visual coding tutorials using learners' annotations [23]. Moore et al. collected crowds' explanations on problems to generate knowledge components [47].

Glassman et al. [22] collected hints from students after they resolved a bug to provide personalized hints to fellow students.

Through learnersourcing, the learners can also pedagogically benefit through the activities designed. Weir et al. asked the learners to generate subgoals from how-to videos which helped the learners grasp the materials [61], and Rittle-Johnson demonstrated the effect of self-explanation and instruction in learning and remembering the learning materials [53]. In addition, Denny et al. have shown the effect of having learners generate multiple choice questions on deep learning of the contents and the final learning outcomes [14]. Specifically in live streams, CatchLive provided catch-up summaries to viewers who join later from both the stream content and the user interaction data [66], and StreamWiki asked viewers to conduct microtasks – writing a summary, commenting, voting – to help other viewers understand the content and context [43].

Following previous adoptions, we applied the concept of learnersourcing in educational live streams. In this work, we aimed to provide interaction cues to instructors through aggregated learners' annotations, while individual learners are instructed to make annotations on top of the live streaming transcript which provides an additional channel for learners to engage with the learning content.

2.4 Using transcripts and transcript-based annotations

Since a major portion of information in educational videos is transmitted by speech, transcripts can be used to capture the majority of the video content. Many systems dealing with educational videos have been built based on transcripts. For example, Dessì et al. proposed an approach to automatically classify collections of educational videos by applying natural language processing techniques on video transcripts [15], and Pavel et al. presented transcripts-based authoring tools that can greatly facilitate authors to create video digests of lecture videos [50]. The success of these designed systems proved the great potential of transcripts in revealing the main content of educational videos. From the perspective of learners' cognition, learning science research showed that transcripts are effective in helping with content comprehension and quick viewing [24, 37, 41]. In addition, Shao et al. claimed that extracting keywords from educational videos can help learners understand and learn terminologies [55], and Fraser et al. presented a temporal segmentation method for creative live stream videos into meaningful segments for the viewers to navigate through [20]. Comparing transcripts with in-video captions, learners prefer using transcripts for following and understanding video content over captions even though transcripts are physically farther from the video than captions, which could be attributed to the longer content history displayed by transcripts [39].

Previous studies also demonstrated the potential for employing transcript-based annotation to support effective interactions. For example, it is evident that transcript-based highlighting can improve group communication quality of multilingual communication [49]. In educational settings, Torre et al. showed that by learning with transcripts augmented with highlighted key concepts in educational videos, learners can achieve better immediate comprehension and increased performance [59]. Liu et al. denoted that active transcript-based annotating helps learners generate better lecture notes [42]. Fang et al. showed that transcript-based annotating can enhance learners' cognitive engagement, and furthermore, displaying transcript-based annotations generated by peers along with educational videos can help with improving learners' sense of learning community [18].

Inspired by previous research that illustrates the potential of transcripts, we employed real-time transcripts and transcript-based annotations in educational live streams in this paper and explored how instructors and learners may benefit from them, especially from their easily reviewable nature.

Participant	*Background*	*Experience*
P1	Graduate student from an American university	Online learning and online teaching experiences
P2	Graduate student from an American university	Online learning and online teaching experiences
P3	Graduate student from a Korean university	Online learning experience
P4	Undergraduate student from an American university	Online learning experience
P5	Undergraduate student from a Korean university	Online learning experience

Table 1. Pilot study participants' information

3 Piloting learners' annotating practices during educational live streams

The instructors' need for learners' interaction cues as a teaching aid is made clear by the literature. We intend to design EduLive as a system that acquires these cues from learners and presents them in the format of aggregated real-time transcript-based annotations to instructors to ensure reviewability during live streams.

However, *how do learners think about the practice of making annotations on transcripts during educational live streams? Will they take notice of the transcripts while watching live streams? Will they actively annotate? Will the learners' annotations be meaningful enough for later use?* While we envision a system that takes learners' annotations as raw "input" to be reorganized into more easily understandable and actionable interaction cues ("output"), it is important to answer these probing questions first before we develop our system, as learners' willingness of performing the practice and appropriately generating annotations are crucial for the envisioned system to function. Without learners' sufficient transcript-based annotations as the input, it is impossible for the system to succeed and produce high-quality cues as a teaching aid as the output.

To understand how learners make use of real-time transcripts when the transcripts become available and generate transcript-based annotations during educational live streams, we ran a pilot study with five participants who are university students.

3.1 Participants

Pilot study participants are recruited through snowball sampling in an American university and a Korean university. Two of the participants were undergraduate students, and three were graduate students. All participants were proficient in English and English was used as the main language of their education programs. They all had extensive experience in online learning for over one year, and two of them had the experience of teaching through educational live streams via Zoom. Table 1 summarizes the participants' information.

3.2 Study setup

In the pilot study, a 38-minute video recording of an introductory-level biology class was provided. Participants were asked to watch the video without controlling the progress bar (i.e., no dragging, pausing, etc.) to simulate the live streaming experience. They were asked to split the screen, with the

YouTube video [6] on the left and the Otter.ai [7] interface on the right. Figure 3 shows a screenshot of the pilot study setup. Through the Otter.ai interface, a real-time transcript of the lecture was shown, and the participants were allowed to highlight and comment on the transcript. The designed setup allowed us to probe learners' practices of making transcript-based annotations while watching live streams. In the pilot study, we provided the tools and features without restricting how learners should use them in order to observe their natural behaviors. A researcher monitored the entire learning process and took observational notes of learners' annotation behaviors (such as what type of content they highlighted). After they finished the study, they were invited to attend a 40-minute semi-structured interview conducted by the researcher to understand how they made use of the transcripts and annotation tools and how they perceived the practice of annotating real-time transcripts in educational live streams. Two pilot study participants were asked additional questions regarding how educators could make use of the annotations for their teaching purposes because they had online teaching experience. The interviews were audio-recorded.

3.3 Analysis

Both interview scripts and observational notes taken by researchers (referred to as "materials") are analyzed following the thematic analysis approach [5]. Two researchers collaborated on conducting an iterative, bottom-up thematic analysis. Initially, each researcher familiarized themselves with the materials by individually reviewing all of them and generated an initial set of codes. Subsequently, they engaged in discussions to combine and form high-level themes. After refining the coding scheme with 22 codes, each researcher independently recoded the materials. Regular meetings were held to review codes and address any discrepancies, leading to the finalization of codes without disagreement. Four themes emerged from this process.

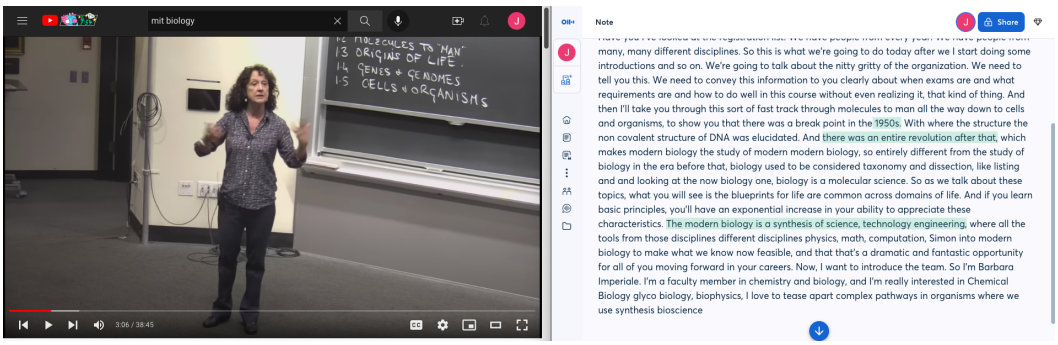

Fig. 3. pilot study setup. Left: A YouTube video. Right: Otter.ai interface that shows the real-time transcript of the video.

3.4 Findings

We present our pilot study findings based on the four themes: (1) opinions about the real-time transcript, (2) use cases and attitudes toward transcript-based highlights, (3) use cases and attitudes toward transcript-based comments, and (4) participants' perceived level of distraction.

Thoughts on real-time transcripts. Most participants (P1, P3, P4, P5) found the real-time transcript to be helpful as long as it is relatively accurate. P1 and P3 mentioned that the transcript

[6]Video source: https://youtu.be/KlVHqq38KJU
[7]https://otter.ai/

could be particularly helpful for people with disabilities or non-native speakers who have difficulties understanding the lecture aurally. P2 used the transcript when she wanted to go back to check a key phrase mentioned in the lecture. Regarding the transcript generated by the specific application used in the pilot study, P5 mentioned that the transcript being updated at the word level can be distracting.

Using highlights. All participants used the highlighting function during video-based learning. Learners tended to highlight the perceived important points (e.g., the definition of terms, keywords), contents that are hard to remember (e.g., numbers, year of a particular event happened), and the parts that they personally felt interesting. For example, P5 highlighted words that she did not understand so that she could look them up after the class. P2 appreciated the easiness of making highlights based on the text, and P4 said that highlights would be helpful when reviewing. When researchers prompted them to imagine the scenario of being able to share these highlights with other learners and the instructor and asked for their opinions, participants expressed interest in looking at others' highlights to check if they missed any important concepts or ideas. For example, P3 mentioned that seeing others' highlights could enhance peer presence, and P5 mentioned that different learners' highlights could work as cues to complement each others' understanding and that the experience of learning with shared highlights may create a similar experience of studying within a group.

Using comments. Three participants used the commenting function (P1, P3, P4). They used this function to correct transcript errors and mark the parts that they were confused about for later review. Some participants expressed concerns about not having enough time to check peers' detailed comments during live steaming, though one participant (P3) expressed that being able to see those comments would provide a sense of learning companion and could be useful after the class. P3 also mentioned that if comments include questions and can be shared with the instructor, the instructor could easily understand the context of the question through the transcript. P4 was concerned about privacy issues and said that she needs to be more careful about wording for impression management when the comments are displayed to the instructor. Both participants who had teaching experience (P1 and P2) worried that they would not have enough bandwidth to read learners' detailed comments when they were teaching, especially when some of the comments were just learners' notes or random thoughts that were not well-framed to be shared with the instructor.

Considerations around multitasking and distraction. Although annotating on the real-time transcript during live streaming did seem to cause minor distraction, most participants (P1, P3, P4, P5) thought it was inevitable, and mentioned that the level of distraction did not differ much compared to existing learning practices such as note-taking during in-person or online learning with pre-recorded videos. Moreover, participants were able to strategically adjust how they use the transcript to lessen the distraction. For example, P3 mentioned that she would choose not to comment or check others' comments when she felt that the instructor was talking about something important that needed to be concentrated on. Overall, learners were more concerned about the distraction caused by checking peers' comments during the live streams, instead of making one's own annotations. The two participants who had teaching experience (P1 and P2) stated that multitasking (i.e., teaching and checking learners' comments) during real-time live streaming may overburden the instructor as the two tasks compete for their attention and cognitive resources.

Overall, participants were positive about employing transcript and transcript-based annotations in educational live streams. We used their feedback, along with previous literature that revealed instructors' struggles and needs during live streaming teaching, as references to design our live stream supporting system, *EduLive*.

4 Design goals

Putting together previous work and findings from the pilot study, we identified that online instructors need a way to "read the class" and assess learners' engagement, while learners in our pilot study were positive about making transcript-based annotations that can be leveraged for these purposes. We spotted a design space for a social computing system that reorganizes the fragmented annotations submitted by individuals and aggregates them into an abstract level that shows learners' collective standing in real time, which can serve as a teaching aid for instructors.

Informed by our understanding of the design space, we identified a list of design goals. It maps the needs and feedback from online instructors and learners to an actionable agenda for system design.

DG1. *Design separate views for instructors and learners.* Instructors and learners have different needs when interacting with the other party and content through an educational live streaming supporting system. For example, learners could benefit from annotation tools for their learning needs while instructors, in general, do not need them; instead, instructors need collective interaction cues and actionable insights, while these are not seen as valuable for learners. The different information needs of the two parties suggest that the system design can benefit from incorporating and optimizing separate views, providing different sets of features to instructors and learners.

DG2. *Minimize distraction induced during the interactions.* Both teaching and learning are cognitively taxing tasks, and instructors and learners would not have much bandwidth for other activities that require significant attention or time. The design of a live stream supporting system should minimize the distraction caused by monitoring or interacting with its interface, avoiding breaking the teaching flow during the live streams. Several strategies can be utilized to achieve this: (1) The annotation interface should be designed to be lightweight, learnable, and accessible. We should only incorporate interface elements that people are already familiar with into the system, instead of introducing new interaction techniques for annotating that require learners to adopt and adapt while learning with the live stream. Our pilot study revealed that learners are concerned about commenting for being unnecessarily time-consuming, and are reluctant to share all comments with the instructors, for example, those that are not being carefully articulated or unrelated to the instructors. Responding to these concerns, we decided to exclude the commenting feature. We employ transcript-based question-asking instead (i.e., ask a question while marking the related part of the transcript) as a method of annotating, as questions are the specific type of message that learners feel comfortable sharing and are directly relevant and informative to the instructors. (2) The frequency of salient visual change on the interface should be minimized. The rapid and salient visual changes on the interface (caused by word-level transcript update) could amplify distraction, as evidenced by the pilot study. In a live stream supporting system, we should consider employing segment-level transcript-update (i.e., appending a new segment to the transcription, usually containing one or two sentences, only when a short period of silence is detected in the instructor's audio) to avoid unnecessary distraction.

DG3. *Aggregate and present learners' annotations in different granularities to support instructors' needs when they are in different situations.* Closely related to ***DG2***, instructors do not have the cognitive bandwidth and time to monitor individual learners' status and actions while actively teaching new content, as documented in the literature [67]. An educational live streaming supporting system should be able to summarize learners' annotations into high-level cues (e.g., summative statistics). This approach can reduce the number of nuanced details that would cause distractions and excessive cognitive processing efforts from the instructors,

enabling the instructors to keep track of them without having to dive deeply into fragmented details. However, some individual annotations may need to be addressed by instructors (e.g., questions), so these details should still be made available to the instructors when needed. The system should reveal details only when the instructors are ready to read. Furthermore, when showing the detailed annotations, the system should format them as actionable items to assist the instructors. These considerations point to the need for incorporating separate modes, in which learners' annotations are aggregated and presented in different granularities (i.e., with different levels of detail).

DG4. *Capture a rich context of annotations.* As pilot study participants mentioned, an important benefit of transcript-based annotations is that they can provide a rich context for instructors, which is especially helpful for understanding and grounding when the instructors come back to questions that learners asked a while ago. To fully take advantage of this benefit in a live stream, we should make sure that each annotation, when being viewed by the instructors, is linked back to its corresponding part of the transcript.

DG5. *Share learners' annotations without increasing the pressure on learners.* A pilot study participant expressed her anxiety caused by impression management (i.e., the desire to leave a good impression on the instructor) and privacy considerations when the annotations are shared with the instructors. This problem can be even more pressing when the learners do not understand what is being shared. Taking these concerns into consideration, a live stream supporting system should aim to give instructors and learners equal access to all shared information. Although such information (e.g., annotations) will be reorganized and displayed differently on the two views (according to ***DG1***), we need to ensure symmetric access to raw information, avoiding cases when one party can access additional information when the other party is not aware of it. Particularly, learners' awareness of what information is made available to the instructors may reduce their anxiety. The system should also hide learners' personal information unless it is relevant for communication purposes (e.g., when an instructor answers a question from a specific learner) [8].

5 EduLive system design

We instantiate the design goals into our educational live streaming supporting system, EduLive. EduLive leverages transcript and transcript-based annotations to re-create interaction cues. It assists instructors in tracking and responding to learners' engagement and confusion. Instead of introducing separate features that achieve the listed design goals one by one, we would like to emphasize EduLive's holistic functionality in transforming the interaction dynamic between the instructors and learners of a live stream, which is achieved by aggregating and displaying learners' annotations in different granularities, taking into account the target users of it (instructors vs. learners) and their current status (in lecture mode vs. in pause mode).

Suggested by ***DG1***, the system includes an **instructor's view** and a **learners' view**. EduLive employs a pausing mechanism, which enables the instructor to switch between (a) the **lecture mode** – focusing on the prepared teaching material while being supplemented with summative information extracted from learners' annotations to keep track of their status and engagement, and (b) the **pause mode** – pausing from teaching new content to read and respond to learners' detailed annotations. This two-mode design is inspired by ***DG3***. The transition between the lecture mode and the pause mode is synchronized between the instructor and the learners.

[8]Learners could not ask questions anonymously in the user study presented in this paper, because the attention check is performed based on the number of questions asked per participant.

5.1 Lecture mode

At the beginning of a live stream, both the instructor and the learners are in lecture mode. In this mode, the instructor focuses on the lecture while being informed of learners' status and engagement on a collective level, which is made available by EduLive's aggregation based on learners' annotations. Everything presented in the lecture mode, in both the instructor's view and the learners' view, stays at an abstract level and cannot be traced back to individuals.

5.1.1 Learners' view in the lecture mode. In the learners' view, the live streaming video and the real-time transcript updated at the segment level (as suggested by ***DG2***) are displayed side by side. A chat box is embedded in the interface and learners can send any messages publicly. Fig. 4 illustrates the learners' view in lecture mode.

Learners are able to make two types of transcript-based annotations while watching the live stream by selecting the corresponding text segment and entering their inputs, namely, *highlighting* and *question-asking*. Guided by ***DG2***, we introduced these annotations because they are (1) light-weight tasks and commonly used features that do not distract learners much from the live streaming video, and (2) both are pedagogically meaningful practices. Learning science research showed that highlighting important parts of learning materials can help learners transfer the information into working memory for further process [60]. Question-asking is the most direct way to clear up one's confusion and is a commonly supported practice in most current live streaming platforms. The learners' own annotations are displayed directly on the transcript, marked with green for highlights and marked with red for questions.

Around the transcript, all learners' annotations are aggregated and visually rendered as graphical bars. Instead of showing the learners their peers' raw annotations (in text format), EduLive shows where learners, as a collective, highlight and ask questions, in a transcript-anchored manner. Highlights are presented in green visualization bars to the left of the transcript, and questions are presented in red visualization bars to the right of the transcript. We adjusted the intensity of the colors (e.g., darker green for more highlights) in the visualization bars to reflect how frequently each part of the transcript is annotated. Visualization bars around transcripts were also adopted in NoteCoStruct, aiming to provide social navigation and enhance learners' sense of community [18].

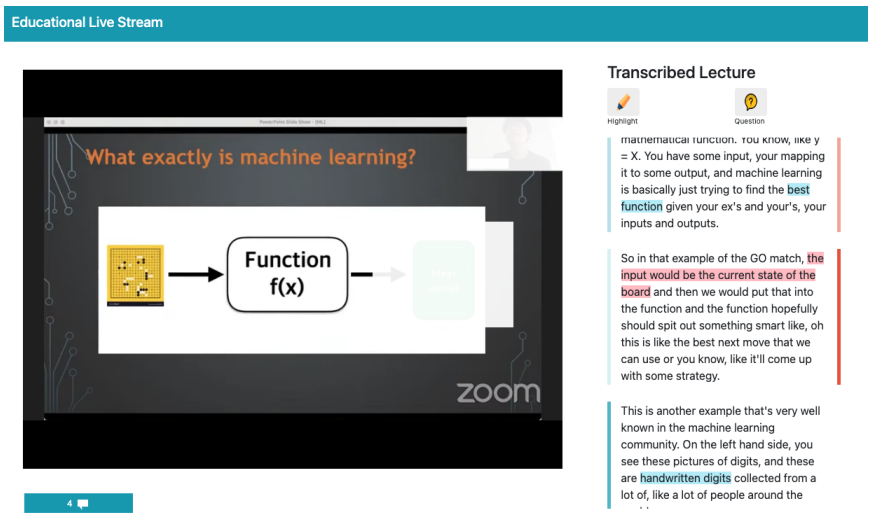

Fig. 4. Learners' view in the lecture mode

5.1.2 Instructor's view in the lecture mode. The three main components of the instructor's view in lecture mode are the *dashboard* of learner engagement and a *real-time transcript* of the lecture, as well as a pause button that enables the instructor to switch to pause mode (see Fig. 5).

The dashboard, which is motivated by ***DG3***, summarizes the learners' annotating behaviors and provides the instructor an opportunity to easily keep track of the class's engagement without being cognitively overwhelmed. Specifically, it shows the number of highlights learners made, the number of questions learners asked, and the percentage of learners who made these annotations (calculated by $\frac{number\ of\ learners\ who\ annotated}{total\ number\ of\ learners}$) since the instructor's last resume time. With the three simple numbers, EduLive provides the instructor a high-level overview of how engaged or confused the learners are in this period of the class, without having to dive deep into all learners' annotations in its raw format (i.e., text). The instructor can refer to the information on the dashboard to strategically navigate the class, such as adapting the lecture pace or content or deciding when to pause the lecture. The real-time transcript is generated from the instructor's audio and is updated at the segment level. Similar to the learners' view, the green and the red visualization bars around the transcript show where learners annotated, while more detailed information (e.g., the content of learners' questions) is hidden during the lecture mode.

Before each live stream, the instructor is asked to set a threshold for the number of questions they would like to collect before pausing the lecture. When the number of questions reaches the set threshold, the "Pause the lecture" button turns from grey to red, alerting the instructor that the threshold has been reached. This automatic alerting feature is also inspired by ***DG3***, aiming at providing directly actionable signals and reducing the instructor's cognitive load for decision-making when lecturing. EduLive uses the number of questions as a criterion for sending pausing notifications because a large number of questions is a sign that learners are confused by recently taught material and have difficulty catching up with the progress, so a pause may be needed. Note that the instructor retains the ability to pause the lecture at any moment, even if the pausing button appears grey. They are not obligated to wait for alerts or pause when alerted.

By clicking on the "Pause the lecture" button, both the instructor's and the learners' interfaces switch to the pause mode of their views.

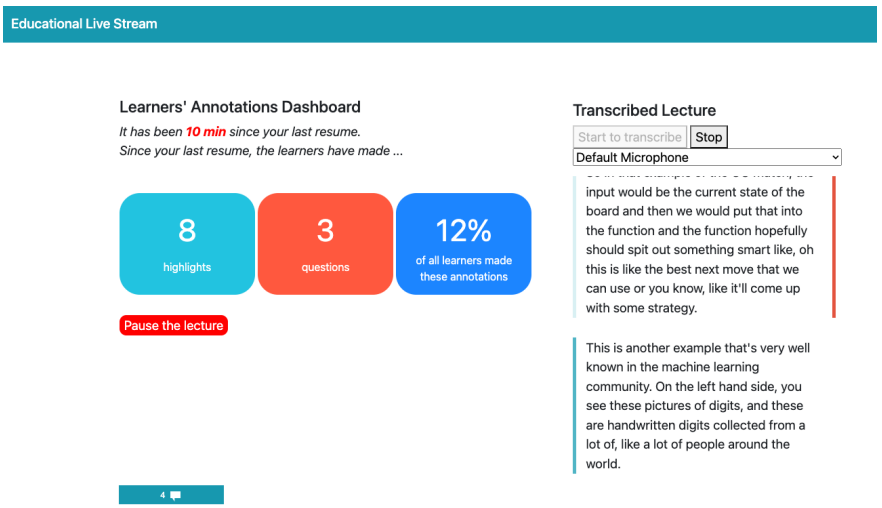

Fig. 5. Instructor's view in the lecture mode

5.2 Pause mode

After the instructor clicks the pause button, both the instructor's and the learners' views are switched to pause mode. In this mode, the instructor actively responds and addresses the learners' annotations. It is worth noting that "pause" is just a metaphor that describes the instructors' and learners' discontinuation of teaching or learning new course material, while the actual live stream (video and audio) and the transcript still go on. In this mode, EduLive guides the instructor and the learners to dive deeper and advises them to address the individual annotations.

5.2.1 Instructor's view in the pause mode. Upon entering the pause mode, an annotation summary box appears below the dashboard displaying learners' detailed annotations (see Fig. 6 and Fig. 7). The summary box is composed of three components: (a) *learners' questions since the last resume*, (b) *previously unanswered questions*, and (c) *top five highlighted segments* since the last resume, or all highlighted segments if learners highlighted less than five. Inspired by systems designed for coordinating crowds (e.g., Apparition [40]), EduLive displays questions in (a) and (b) in two checklists. This design decision also follows ***DG3***, which suggests that annotations should be displayed as action items so that the instructors can handle them more easily. In addition, the checklists not only help the instructor keep track of which questions were answered and which were not but also synchronize learners with the instructor's current progress.

Zooming in on the questions in the summary box (see Fig. 7), in both (a) and(b), each checklist item includes a learner's question in full (text in black) and the associated part of the transcript (text in grey) which is selected by the learner when they make the annotation. The question asker's name is displayed, in case the instructor deems it helpful for communication. Questions based on the same segment of the transcript are merged into one question block. These questions are likely to be highly related and thus may be answered together. Question blocks are ranked based on the number of questions they include, from highest to lowest.

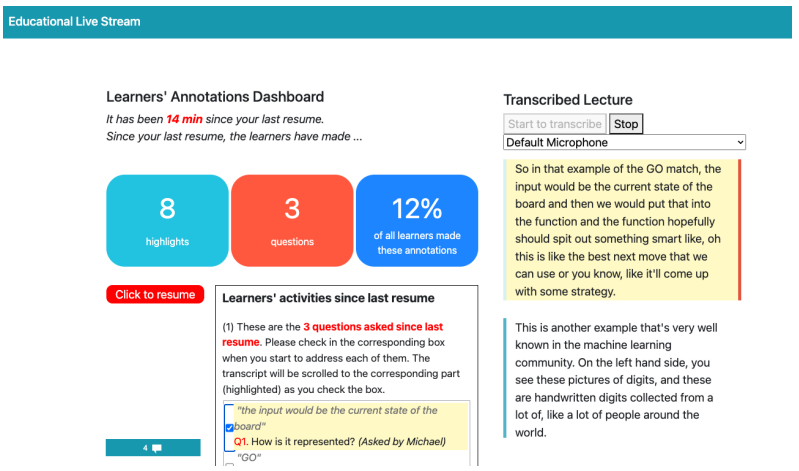

Fig. 6. Instructor's view in the pause mode

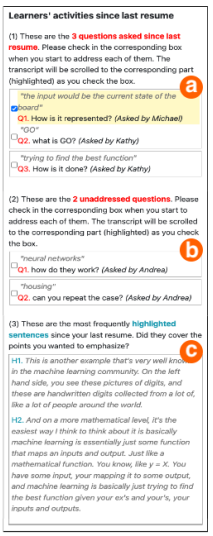

Fig. 7. The complete summary box in the instructor's view

To answer a question, the instructor checks its corresponding checkbox. Then, the transcript is automatically scrolled back to the segment that the question is based on, with the segment highlighted. In this way, we make sure that the broader context of each question is not lost and the benefit of the transcript-based annotating method is preserved, which achieves ***DG4***.

By clicking on the resume button, both the instructor's and the learners' views are switched back to lecture mode.

5.2.2 Learners' view in the pause mode. In the learners' view, the size of the video frame is reduced, and learners' attention is directed to the summary box that appears below the video (see Fig. 8). Guided by ***DG5***, we maintain the symmetry in information accessibility by containing the exact same set of information in the summary boxes on the learners' view and the instructor's view. The summary box on the learners' view also contains the (a), (b), and (c) components as shown in the instructor's view (showing the detailed annotations made by peers) introduced in the instructor's view, except that the learners cannot control the checkbox. When the instructor starts to address a particular question by checking the checklist item, the auto-scrolling and highlighting of the real-time transcript will take place in the learners' view as well. Through the synchronized auto-highlighting and auto-scrolling between the instructor's view and the learners' view, EduLive effectively directs the learners' attention and makes sure both parties are on the same page during the pausing time.

Learners' view is switched back to the lecture mode when the instructor resumes.

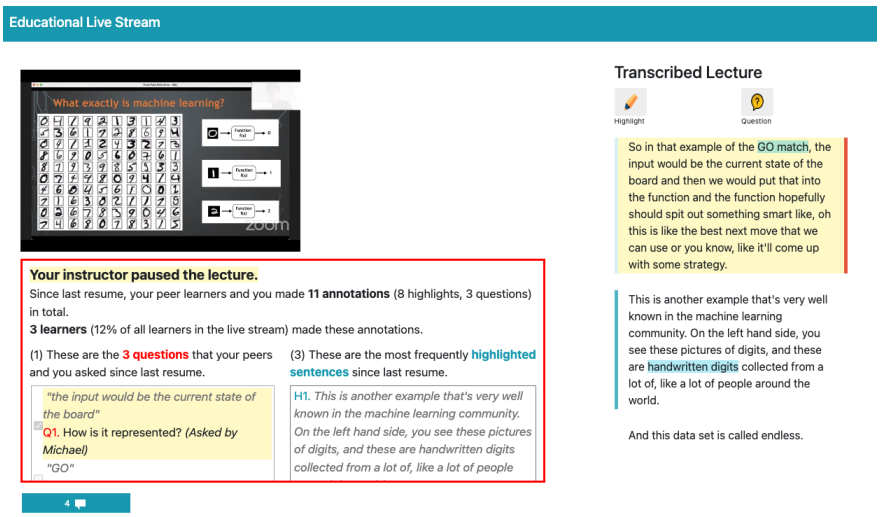

Fig. 8. Learners' view in the pause mode

5.3 Implementation

Both the instructor's view and the learners' view were implemented with HTML/CSS and Javascript. We used the Firebase real-time database to sync the information between the two views.

The real-time transcript was generated using Microsoft Cognitive Services Speech SDK [1]. The speech-to-text algorithm segments the transcript automatically based on the length of silence in the audio, thus each transcript segment may contain one or more sentences. The service has several

technical limitations, for example, the accuracy of the transcription could be affected by acoustic quality, non-speech noise, etc. [9].

We chose to update the transcript at the segment level even though word-level updates are technically achievable and can reduce latency for two reasons, (a) according to DG2, we should avoid the transcript being constantly updated, which escalates distraction, and (b) the word being recognized immediately after the instructor speaks it is not finalized and may be changed based on the context, and we wanted to prevent learners from annotating on the text that may disappear later.

6 Evaluation method

We deployed EduLive in four live streaming classes that involved instructors and learners, comparing EduLive to two alternative designs. With four instructors giving lessons on unrelated topics and having vastly different teaching styles and preferences, yet are particularly representative of real-world teaching scenarios, these four classes serve as a proof-of-concept evaluation of the design of EduLive and gave us a starting point for investigating how transcript-based annotations can impact the dynamics of educational live streams in practice in sessions with 10s of audiences. Similar approaches and user study scales have been used in previous research for evaluating live streaming tools and features (e.g., [44, 65].

In this section, we first describe the design of the two baseline systems, EduChat and EduScript. We also detail the process of recruitment and data collection.

6.1 Baseline systems

6.1.1 EduChat. EduChat is the simplest baseline system that enables learners and the instructor to communicate through the chat box. It is designed similarly to popular live streaming platforms such as YouTube Live and Twitch. Specifically, the live streaming video and the chat box are incorporated into the learners' view, while the instructor's view only includes the chat box.

6.1.2 EduScript. EduScript is another baseline system that employs a chat box as the only communication channel. On the basis of EduChat, the real-time transcript is added to both the instructor's view and the learners' view. In addition, the learners are able to make transcript-based annotations (i.e., highlight and ask questions) in the same manner as in EduLive, but the annotations are for individual use only and are not shared with other learners and the instructor in any format (e.g., no visualization bars, dashboard, etc.).

6.2 Participants

6.2.1 Instructors recruitment. We recruited four instructors (two from an American university, and two from a Korean university) to use the three designed prototypes to conduct educational live streams through social media. We invited instructors who have at least one year of previous in-person or online teaching experience, and have no speaking or reading disabilities or colorblindness. We introduce their backgrounds and teaching experiences in detail.

The instructor of basics of machine learning (I-ML) has three years of teaching experience in STEM subjects. Most of his classes are three-hour-long sessions (including approximately two-hour long lectures and one additional hour for interactive activities with students) for middle school students. He initially started with teaching in-person classes, but later shifted to teaching through online live streaming using Zoom due to the COVID-19 pandemic.

[9]https://learn.microsoft.com/en-us/legal/cognitive-services/speech-service/speech-to-text/transparency-note. Accessed in 2023.

Instructor	*Background*	*Topic*	*Online Teaching Experience*
I-ML	Course instructor from a Korean University	Machine Learning	About 2 years
I-CC	Teaching assistant from an American university	Chinese Culture	About 2 years
I-GW	Course instructor from a Korean university	Global Warming	About 2 years
I-SC	Course instructor from an American university	Sociology	About 1 year

Table 2. Instrutors' information

The instructor of Chinese culture (I-CC) has been working as a teaching assistant for four years at a university. I-CC also has both in-person teaching experience and online teaching experience through Zoom. I-CC chose to teach topics around Chinese culture because of her personal interests and expertise, as well as her previous experience in teaching and presenting related topics at multicultural events.

The instructor of global warming (I-GW) has about 10 years of teaching experience in STEM-related subjects, with 2 years of online teaching experience. His target audiences are mostly K12 students. He taught three-hour-long discussion-style or experiment-style classes and tried to keep classes interactive to avoid learners losing interest or being distracted during the long classes.

The instructor of introduction to sociology (I-SC) has six years of teaching experience in college-level classes. She has extensive experience in various styles of classes, from main lectures to discussion sections and design studios. I-SC taught both in-person classes and online live streaming classes through Zoom.

Table 2 summarizes the instructors' information.

After each instructor agreed to attend the study, a researcher scheduled a 50-minute long meeting with each of the instructors to give an overview of the study, describe the required experiment setup, walk the instructor through the three prototypes that they will use during their live streaming sessions, and provide a guideline of how to prepare for the live streams (e.g., recommended font size of the slides, expected time length, etc.). Afterward, the instructors each scheduled three 30-minute time slots for the live streams and a 50-minute time slot for the semi-structured interview within 2-3 consecutive days. They were asked to deliver the live streams in a quiet environment to enhance acoustic quality, thereby boosting the accuracy of the transcripts. To avoid instructors' fatigue, the sessions are scheduled for at least 25 minutes apart from each other.

The instructors were asked to pick a topic that they were highly familiar with and did not require prior knowledge to understand, prepare the slides, and send them to a researcher at least two days before the first live streaming session. A researcher would contact the instructor if the content needs to be adjusted (e.g., if a certain part is conceptually challenging for people who do not have prior knowledge).

After finishing all three live streaming sessions, the instructor was interviewed by a researcher. Each instructor was compensated with a $50 Amazon gift card upon completion of the study.

6.2.2 Learners recruitment. We posted a screening survey as a task on Prolific [10] for each live streaming session. We sent email invitations to respondents who are interested in learning about

[10]https://www.prolific.co/

the topic of the live stream but do not have much previous knowledge about it and can join and concentrate on the educational live stream at the scheduled time. For each session, we invited approximately 25 learners estimating a 50% participation rate, though the actual number of learners who attended the live streams and finished the studies varied from 5 to 15 ($M = 9.33, SD = 2.53$). The learners were asked to learn through an educational live stream and complete a post-survey. The task took about 40 minutes in total to complete. Learners who passed the attention checks were paid $8 through the recruiting platform. We checked if they watched the live streams for at least 20 minutes and interacted as required in the study instruction, and left reasonable responses in the post-survey (e.g., answered factual questions about the interface correctly).

6.3 Study procedures

In each class, one instructor conducted three live streams with the same content, each using a different prototype – EduLive, EduScript, or EduChat. To avoid the potential ordering effect, the orders of prototypes being used by the four instructors were randomized.

The overall process of all live streaming sessions was the same. At the beginning of each session, the instructor joined a Zoom meeting room initiated by a moderator (acted by a researcher of the study). The moderator then streamed the Zoom meeting onto YouTube Live, and the live video was embedded into the corresponding prototype that the learners were using. We decided to let the moderator stream the Zoom meeting to YouTube Live instead of letting the instructor host live streams on YouTube Live directly because the instructors needed to screen-share their slides, which can be easily done on Zoom. While broadcasting software such as OBS [2] could be used to screen-share and stream to Youtube Live and is widely adopted by experienced streamers, our recruited instructors were not familiar with this practice. After sharing their slides, the instructor was asked to log in to the assigned study prototype for the session. We asked the instructor to display the slides and our prototype interface on two separate screens so that both interfaces could be seen at all times during the live stream. The instructor started teaching at the scheduled time, and each live stream lasted about 30 minutes long.

The learners were asked to log in to the prototype website at least 5 minutes before the live stream start time to carefully read the consent form and instructions. During the live streams, they were expected to focus on learning from the live stream at all times and interact with the prototype. We asked each learner to ask at least two relevant questions during the session as an attention check. After the live streams, they were directed to the post-survey page.

In each live stream, the moderator took observational notes that facilitated the later interview with the instructor without interfering with the live streaming content and interactions.

6.4 Measurement and analysis

A researcher conducted a semi-structured interview with each of the four instructors after they finished all three live streams. The instructors were asked about their teaching experience using each of the prototypes and their prior teaching experience. The interviews were audio recorded and transcribed into text. Two researchers independently open-coded the transcripts of the interviews and had a discussion to resolve major discrepancies and reach a consensus on the interpretation.

We measured learners' learning experience with the prototypes using surveys, focusing on the following themes: co-presence with other learners, easiness of participating and asking questions, easiness of sharing and expressing to peers, understanding of questions and answers in the live stream, reviewability of the live stream, sense of agency, and self-efficacy of learning. Some of these themes were adopted because they are often used in learning science or HCI studies for measuring user experience with new interactive systems (e.g., [16, 52]). We additionally added several themes particularly relevant to our study (e.g., reviewability of the live stream and easiness of participating

and asking questions) as we hypothesized that the functionality of EduLive could help improve these aspects. All survey items were presented on a 7-point Likert scale, in which 1 represents "strongly disagree" and 7 represents "strongly agree". We also asked open-ended questions to collect learners' detailed feedback. The survey questions are attached in the Appendix.

We applied one-way ANOVA to analyze learners' quantitative results and qualitatively coded learners' responses to open-ended questions.

7 Result

In this section, we present our findings from the interviews with the instructors of the four classes. We also summarize the learners' experiences with the three prototypes, probed using the post-study survey.

7.1 Instructors' experience

7.1.1 Benefits of employing real-time transcripts and transcript-based annotations during educational live streams.

A number of themes pertinent to the benefits of teaching with real-time transcript and transcript-based annotations emerged in our analysis.

The context provided by the transcript helped both the instructor and the learners. All instructors mentioned that the transcript in EduLive was helpful for getting the context of the questions. I-ML appreciated that transcript-based questions aided him in better understanding and answering the questions. For example, when using EduLive, there was a learner's question that confused I-ML for a few seconds. However, he commented, *"The fact that the transcript actually highlights the part and goes to the part that they used to ask the question helped me understand the context behind why they asked the question"*. Similarly, when using EduScript, a learner asked a question that I-CC could not understand because she did not know what the question was referring to. Although the transcript was provided in the instructor's view of EduScript, I-CC did not want to take the time to check back in the long transcript. Yet, when using EduLive, I-CC said that *"Even though the transcript may make some mistakes, I can see where their origin is, then I [can] understand what's the student's question. ... I just click [in the checkboxes of the] questions and go back to where [the students] marked in the transcript, ..."*. I-GW also stated that *"I [could] quickly identify where the participants got their questions from. ... it's very easy to track the sources of the questions"*.

The dashboard in EduLive helped with assessing learners' presence and engagement. The dashboard in EduLive shows real-time summative information, which helped the instructor get a sense of learners' presence and engagement at a glance. I-ML commented that *"The problem of [EduScript and EduChat] is that when you are checking the questions, you cannot look [on a] high-level or abstract-level. Instead, you have to look into the exact text to justify whether there is a new question or lot, while [EduLive] made it clear"*. When using EduLive, the fact that I-ML could "glance at the interface" to see if he should continue with teaching made him less distracted and led a to smoother live streaming experience. In addition, I-ML felt that he was more aware of learners' presence during the teaching process due to the existence of the dashboard, whereas in contrast, he said that *"[in EduScript and EduChat] I don't know what the [learners] are doing. It feels sort of empty"*. When facing a group of learners, their collective engagement could be more meaningful for the instructor than the engagement of each individual. I-GW indicated that when learners contribute to the chat in EduChat and EduScript, he *"couldn't identify who is who. But it could have been one person speaking the whole time, and one person not speaking"*. Yet, when using EduLive, *"[It shows] the percentage of all learners who made these annotations, and these numbers are of great help"*. I-CC had a similar thought: *"[When using EduChat and EduScript,] I didn't check who asked the questions,*

so it might be one person kept asking. ... I would consider the students, in general, are more engaging when more of them asked questions".

In addition, I-SC commended that EduLive showing learners' engagement in an intuitive way boosted her confidence when teaching. I-SC commented, *"I see that the number of engagements is increasing [on the dashboard], and I became more confident ... maybe [the topic I was talking about] is something interesting for [the learners]".* I-GW hinted that the easiness of making annotations on the learners' side contributed to his increased awareness of learners' engagement. He commented, *"The chat doesn't mean anything unless they actually type in and press enter. But [for making] the annotations all they have to do is press a button, so I think it's a lot easier for them to express their thoughts through highlights than [through] the chat".*

Pause notifications in EduLive were helpful for mode-switching, and the pausing mode provided opportunities for making adjustments during the class. The pause notification, which was activated after the total number of questions reached the instructor-set threshold, helped I-ML decide when to pause the lecture to address the questions. I-ML said that *"[When using EduChat and EduScript] I always have to check the chat box. ... I would always have to read the latest message and see whether the question is [a new one]. With [EduLive], ... it changes the color [of the pause button] for you, so you automatically know that there's a new question. I guess it saved a little bit of time and a little bit of the headspace".* He additionally commented that having the notification show up on the button helped him switch to the pause mode easily.

I-SC commented that the pause notification reminded her to pause once in a while instead of leaving all questions until the end of the class. As she saw the learners' annotations promptly during the class, she was able to make real-time adjustments in the lecturing pace and content accordingly during the live stream. In the session with EduLive, I-SC expanded more on a subtopic as she saw learners' extensive interests in a keyword she mentioned. The pausing time gave her a chance to carefully process learners' annotations and organize her thoughts on how to make adjustments for the remaining class. However, when using EduChat and EduScript, I-SC paused only at the end of the class, which left her no space for further adjusting.

The organization of questions in EduLive's summary box helped with strategizing question-answering. Of the three live streams that I-ML hosted, the session that used EduLive had the most number of learners and questions. I-ML actively compared two different interfaces, EduLive and EduScript, and acknowledged that the way EduLive organizes the questions (as checklist items) helped him to strategize question-answering. I-ML said that *"[If] there were as many people in the second session [using EduScript] as there were in the first session [using EduLive], and I also got as many questions in the second session, I would have been very lost. ... It would have been hard to address all of those questions with the second interface. ... I did enjoy from the first interface that the organization part helped me strategize what questions I can answer".* I-ML also mentioned that the question blocks in the summary box – the clustering of the questions asked based on the same transcript segment – helped him to organize his thoughts and avoid answering similar questions repeatedly. I-ML said that *"For the [sessions using EduChat and EduScript], ... I would just address the questions [in the chat box] from top to bottom. I do feel that method made me repeat myself a bit because people would ask similar questions that could be addressed in similar ways.".* I-SC mentioned that the questions are displayed in a more structured way in EduLive (rather than just a list of questions in a random order in the chatbox), and gave her a better sense of which parts are learners more interested in or confused about on a high level. The strategically organized question list further assisted her in making adjustments during the live streams.

7.1.2 Teaching styles, conditions, and preferences affect how instructors use and perceive transcript-based annotations.

Usefulness of certain features may depend on the duration of live streams. I-ML commented that he would use EduLive more effectively for longer online classes to check the engagement level of the students. Based on his experience of teaching three-hour online classes, in which *"lecture material tend to be less interactive and any physical activity was basically impossible"*, he often felt like the students were less engaged and easily lost focus compared to face-to-face classes. I-ML said, *"Having the information of how my students were engaging with the material would have helped me a lot in those two years (of online teaching)"*. In contrast, he mentioned, *"If I had to teach with [EduChat or EduScript] the whole time, I would have felt like I was talking to myself or talking to a wall. ... Even if there is a chat box, ... there are always going to be periods of time when no one has any questions, so during that time I would feel unsure whether the listeners were actually listening or not. ... [In a long class] the dashboard would give way more information that would be relevant in that class in that situation"*.

I-ML also mentioned that he was not able to fully experience the functions provided in EduLive due to the limited time in our study. He did not look into learners' aggregated highlights during the study, but he stated that for longer live streams he would have used learners' highlighting information and even made adjustments to the content depending on the learners' engagement with the lecture. He said, *"If I had more time I could have even gone through the transcript and look through what parts that they're engaging with the most. ... And see oh, they really engaged with this particular subfield of this subject that I was covering, then I could go into more details on the specific [subfield] instead of something else that they would not have enjoyed as much"*. I-SC said that she might consider asking learners to use the pausing mode as extra time for entering questions and annotations in longer classes, but she was not able to do so in the short class due to the fast pace of our user study.

The strength of employing transcript-based annotations is limited in interest-group-style classes, in which the instructors expect high-frequency back-and-forth interactions. I-CC introduced a topic of her personal interest in her class and expected substantial open discussions. During all three live streams, I-CC highly encouraged learners' interactions and reacted as soon as possible in all three live streaming sessions, leading to interest-group-style sessions with high-frequency back-and-forth interactions. Although I-CC pointed out several advantages of EduLive, some of the features of EduLive were not perceived as relevant and useful as we conceptualized because of I-CC's personal teaching style and preference. Compared to questions, I-CC thought learners' highlights were less relevant to her teaching process, so she did not make use of related information (e.g., the number of highlights in the dashboard, and the visualization bars that reflect where the learners highlighted). Also, since I-CC paid close attention to the chat box messages at all times during the three live streams and tried to respond to each new question/message immediately after it popped up, she did not perceive EduLive to be especially helpful for enhancing her awareness of learners' presence (compared to EduChat and EduScript). Although I-CC was only able to finish 2/3 of the content she prepared when using EduChat since she was frequently interrupted by new questions that appeared in the chat box, she enjoyed spending time on the interactions and did not feel rushed to finish all the slides.

Real-time transcript could be a double-edged sword, and it should be carefully integrated into live stream supporting systems to avoid potential negative effects. Being able to see a transcript generated in real-time during a live stream is not a common practice and people are not used to it yet. I-GW mentioned, *"I thought that was very interesting and kind of scary at the same time seeing that being done automatically. It was very interesting and very hard to adjust at first. ... my mind is sort of subconsciously looking at the transcription as well at the same time [while teaching]. I think that's a downside of having that transcribing system right next to you"*. The errors in the transcript may also elevate instructors' pressure. To avoid the transcript competing for

the instructor's attention, a more careful design when integrating the transcript into live stream supporting systems should be deliberated. For example, the negative effect I-GW experienced may be mitigated by making the transcript less visually obvious to the streamer. Future designers can consider only making the text available on the instructor's side when the instructor has the intention to check it (e.g., by hovering over a specific place on the interface).

Table 3 summarizes the key benefits and use cases afforded by features of EduLive.

Features provided by EduLive	Benefits and use cases
The reviewability of the transcript and the auto-scrolling function when addressing questions.	(1) Helping instructors to better understand the questions so that they can handle them properly. (2) Enabling instructors to trace back to the corresponding context when addressing questions, which is especially valuable when the questions are not well-framed (3) Could be archived as a resource for future course preparation.
The dashboard that shows summative information extracted from learners' annotations.	(1) Supplementing awareness of learners' presence and providing instructors cues to sense whether learners are actively engaging on a group level. (2) Boosting instructors' confidence in teaching.
The pausing mechanism that separates live streams into lecture mode and pause mode.	(1) Avoiding distraction caused by frequently going back and forth between the slides and the chat for checking new questions, which is the default practice when a pausing mechanism is absent. (2) Providing opportunities for making adjustments to teaching pace and content.
Summary box that strategically organizes learners' questions.	(1) Helpful for managing a large number of questions and avoiding answering similar questions repeatedly. (2) Assisting with making adjustments to teaching during live streams.

Table 3. Summary of key benefits of EduLive perceived by instructors.

7.2 Learners' experience

Since in all four classes, the number of learners who participated was relatively small, and the number of learners in each session varied (e.g., in Class 1, there were 5 learners using EduChat and 9 learners in EduLive), we merged the learners of the four classes into three groups based on the prototypes they used in the analysis. This yielded relatively balanced samples of the three conditions for statistical comparison ($n_{\text{EduChat}} = 38, n_{\text{EduScript}} = 34, n_{\text{EduLive}} = 40$).

7.2.1 Quantitative results. Learners actively made annotations during all sessions. Table 4 shows the number of annotations learners made during each session [11].

ANOVA showed that learners using EduScript and EduLive perceived significantly higher reviewability of the live streams than learners using EduChat, $F[2, 109] = 24.9$, $p < 0.01$ (EduChat: $M = 3.17, SD = 1.85$; EduScript: $M = 5.40, SD = 1.61$; EduLive: $M = 5.63, SD = 1.55$). Post-hoc analysis showed that there is no significant difference in perceived reviewability between learners who attended the four classes, which confirmed that the reviewability of live streams was not affected by the live stream topic. The interaction effect between the system used and the live streaming topic is also not significant. Reviewability of the live streaming content is important for

[11] In sessions conducted with EduChat, no highlight was made as the feature was not provided.

Session	No. of annotations	Session	No. of annotations
S1, *L38 — L46*, w/ EduLive, $n_{\text{learners}} = 9$	40 questions, 68 highlights	S7, *L58 — L72*, w/ EduLive, $n_{\text{learners}} = 15$	59 questions, 137 highlights
S2, *L24 — L30*, w/ EduScript, $n_{\text{learners}} = 7$	15 questions, 73 highlights	S8, *L48 — L58*, w/ EduChat, $n_{\text{learners}} = 11$	34 questions
S3, *L12 — L16*, w/ EduChat, $n_{\text{learners}} = 5$	19 questions	S9, *L73 — L82*, w/ EduScript, $n_{\text{learners}} = 10$	46 questions, 168 highlights
S4, *L1 — L11*, w/ EduChat, $n_{\text{learners}} = 11$	29 questions	S10, *L103 — L112*, w/ EduScript, $n_{\text{learners}} = 10$	46 questions, 235 highlights
S5, *L17 — L23*, w/ EduScript, $n_{\text{learners}} = 7$	24 questions, 103 highlights	S11, *L92 — L102*, w/ EduChat, $n_{\text{learners}} = 11$	35 questions
S6, *L31 — L37*, w/ EduLive, $n_{\text{learners}} = 7$	20 questions, 42 highlights	S12, *L83 — L91*, w/ EduLive, $n_{\text{learners}} = 9$	25 questions, 92 highlights

Table 4. Summary of the number of annotations learners made in each live streaming session.

learners in two ways. First, reviewability enables learners to make annotations based on the content mentioned by the instructor a while ago. Second, when the instructor is answering a question asked by a learner previously, reviewability enables learners to trace back to understand the context of what the instructor is answering so that they can be on the same page as the instructor.

ANOVA did not reveal significant differences between learners using the three prototypes in co-presence with other learners, easiness of participating and asking questions, easiness of sharing and expressing to peers, understanding of questions and answers in the live stream, sense of agency, and self-efficacy of learning (all $p > 0.05$).

7.2.2 Qualitative results. Based on the responses to the open-ended questions, many learners using EduScript and EduLive found the transcript to be helpful. Several themes that emerged from learners' responses are as follows.

Transcripts augmented learners' learning ability and helped with formulating questions. The transcript provided additional modality to the live stream and could be helpful for learners who prefer learning through text instead of video and audio. For example, L33 mentioned that *"I learn better from reading than listening, so it was really helpful to be able to read"*. The reviewability of the transcript was also highlighted by learners. L20 stated, *"I used it to read back information from earlier"*. Similarly, L18 mentioned that *"It helped me to follow along and to [go] back to look at any information I wanted to look at again"*. Learners also stated that the transcript helped them formulate questions. For example, L40 mentioned that *"I used it to clarify certain points that I didn't immediately understand or confirm what I thought was said. It helped me understand better and helped me find where/how to ask good questions"*. Similarly, L41 mentioned that *"It was helpful to be able to read the transcript when I wanted to ask a question so I would be able to word the question properly"*. However, some learners also pointed out that the latency and inaccuracy of the transcript made it less useful than expected. For example, L45 mentioned that *"I wanted to use it more than I did. I like*

to read along with a lecture because I learn better that way, but the transcript was too slow for that". L22 mentioned that *"... it didn't perfectly record the words, so some weren't accurate to what was said".* We discuss considerations around transcript inaccuracy and latency in the Discussion section.

Annotation summary box helped learners to keep track of questions and answers. Many learners using EduLive perceived the annotation summary box that appeared during the instructor's pauses to be helpful. Specifically, they benefited from the synchronized questions checklist. For example, L87 mentioned that *"... it was easy to follow along when the instructor was answering the questions.".* Similarly, L38 reflected, *"It made it very easy to follow what classmates had asked and anticipate what would be said so it was quite useful.".* L41 commented, *"It was really good to be able to see exactly which questions the streamer was answering. Being able to read it alongside the streamer while he was addressing them made it much easier to understand".* Many learners enjoyed reading peers' questions in the summary box during the pauses and also tried to avoid asking similar questions to reduce the instructors' burden. L45 stated, *"I looked at the questions other people were asking so I could avoid asking a similar question".*

Learners had different opinions regarding visualization bars around the real-time transcripts. Some learners found the green and red visualization bars around the transcripts to be useful. For example, L68 mentioned, *"[With visualization bars] I could see when others had made notes or questions during a certain part and could tell when a seemingly important point or puzzling part came up, and that others noticed it too.".* While some learners complained that *"almost everything is a little green"* which dampened the usefulness of the visualization bars (L89), others commented that *"someone was just highlighting everything which wasn't very helpful, but the darker green indicated more important passages"* (L90). We also noticed that a considerable amount of learners did not pay much attention to the visualization bars because they were mainly focusing on the fast-paced live streams. Some learners commented on how visualization bars could be more beneficial after the live streams than during the live streams. For example, L65 stated, *"The transcript moved too fast for me to use them, but I think I would use them if I were reviewing the material– for instance, to make sure I hadn't missed something important that many other students had put a highlighter on".*

In general, learners using EduLive were positive about their experience. L38 commented that the system was *"so good at enabling people to communicate meaningfully".* Similarly, L41 mentioned that *"Overall I think it was a good interface and it made it easy to follow along and understand what the streamer was teaching".* Several learners also mentioned that they might be more comfortable with using the system after they get used to it. For example, L35 mentioned that *"It's good to have options [to look at on the interface]. But more options means more of a learning curve. I would feel more comfortable with the system after a bit more practice".*

8 Discussion

Our results from the evaluation suggest that EduLive, as an educational live streaming support to re-create interaction cues, could reshape the ways instructors and learners interact in educational live streams. Instructors and learners also experienced unique benefits from transcript-based annotations being strategically aggregated in two modes. In this section, we discuss the lessons we learned from designing and evaluating EduLive, reflect on the limitations of our study, and shed light on future work in this design space.

8.1 How transcripts and aggregated transcript-based annotations could affect instructors' teaching experience

Since teaching experiences, subjects to teach, and personal preferences can all shape an individual's teaching style (e.g., lecture-based versus discussion-based) adopted in the class, the features and mechanisms provided by EduLive may become more feasible in some scenarios than in others.

On account of instructors' contrasting teaching styles and subjects, we are provided with an opportunity to compare their experiences and feedback to conclude when and how can transcript and transcript-based annotations support instructors' needs in educational live streams.

Instructors identified several common benefits of utilizing EduLive. ***In response to RQ1***, we illustrated that learners' transcript-based annotations, after being aggregated and restructured by EduLive, could reshape the class process and dynamics, provide cues that effectively raise instructors' awareness of learners' status, and scaffold instructor-learners interaction, as shown in Table 3.

In response to RQ2, we found that instructors teaching in different styles use EduLive differently. Notably, instructors' expectations on the level of interactivity of the class shape much of how they utilize the transcripts and aggregated annotations. Drawing on insights from instructors' feedback, we anticipate that EduLive's components and functionalities hold greater value in lecture-style live streams than in discussion-style live streams. There are two possible reasons behind this.

The instructor's attention and time management, facilitated by EduLive's pausing mechanism, are more relevant in lecture-style live streams. Instructors hosting lecture-style live streams make a serious effort to deliver the content they prepared completely and thoroughly, which requires them to stay focused on their planned teaching and carefully control the time they spend interacting with the learners. Therefore, they are more likely to appreciate the benefits of abstract information (shown on EduLive's dashboard) extracted from learners' overly detailed annotations when lecturing, as well as the pausing mechanism that augments their ability to manage time spent on lecturing versus time spent on responding to learners. In comparison, discussion-style live streams usually require a high level of real-timeliness of interactions, which might be weakened by the pausing mechanism.

The on-topicness of the instructor-learner interaction, enhanced by the transcript-based nature of annotations in EduLive, is more relevant in lecture-style live streams. As annotations in EduLive are generated based on parts of the live stream transcript, they are more likely to be on-topic than freely generated chat messages, which may help the instructor stay focused and stick to the main teaching content when interacting with learners. However, as some discussion-style live streams are intentionally designed to be led by learners' thoughts and interests, the benefits of the transcript-based annotations' on-topicness might be lessened.

In general, while EduLive's design components and functionalities bring benefits, employing aggregated transcript-based annotations is not the optimal solution for supporting all kinds of educational live streams. System design should be tailored according to specific instructor's needs and expected interaction mode. Future work can consider designing more customizable live stream supporting systems and leaving space for instructors to adapt based on the specific scenarios of their classes.

8.2 Learners' experience with transcripts and annotations-driven interactions during educational live streams

Learners overall held positive attitudes toward the transcripts and transcript-based annotations enabled by EduLive. ***In response to RQ2***, learners benefited from the affordances of transcripts, especially their reviewability, to augment their learning and interaction abilities. It is especially encouraging to see that, during instructors' pauses, learners perceived the synchronized questions checklists and transcripts auto-scrolling to be helpful in keeping up with questions and answers and enhancing their comprehension. In other words, the changing interface elements (e.g., highlighted questions currently being answered, scrolled and highlighted transcript showing the context of the questions) on the learners' view in accordance with the instructors' actions at the moment magnified the level of synchronicity of the live stream, which nudges learners to keep pace with

the instructor cognitively. Research in educational psychology also indicated that synchronicity has positive effects on learners' satisfaction, interest in course content, sense of belonging, and cognitive commitment and contribution [9, 17, 30, 46, 51]. The design of EduLive presents a step towards strengthening the synchronicity between the instructor and the learners, on top of the synchronized nature of live streaming videos, by the dynamically changing interface design in response to both parties' activities.

Although some learners thought the latency and inaccuracy of real-time transcripts dampened the usefulness of related functions, it appears that many learners still benefited from the imperfect transcripts. As the reliability of speech-to-text technologies is improving over time, we expect transcript and transcript-based annotations employed in educational live streams could benefit learners even more in the near future.

8.3 Implication to employing transcript-based annotations in longer sessions and at larger scales

Although the time length of each live streaming session was relatively short and the size of audiences was small in the four simulated classes, participants' reactions and feedback provide us with a starting point for considering the additional effects of transcript-based annotations when they are used in longer sessions and in larger classes. For example, when learners are more cognitively drained in longer sessions, the EduLive dashboard could make learners' inattentiveness observable to the instructor so that adjustments to the teaching pace or tone can be made. Furthermore, when the instructors are given more time flexibility in longer sessions, they also have more opportunities to interact with the transcript-based annotations. In particular, instructors could look into learners' highlights and speculate whether learners grasped the key messages, which is not realistic in the short classes, as implied by I-ML during the interview. Similarly, the benefit of aggregating transcript-based annotations may also increase when there are more learners (i.e., greater collective brainpower) annotating the learning material. For instance, compared to a small group of learners each highlighting different sentences, the aggregation of highlights made by a larger group of learners is more likely to collectively identify the key takeaways of the live stream. Hence, beyond the scale of the four classes in our study, transcript-based annotations may offer greater advantages to instructors and learners in longer live streams with larger groups of audiences as the burden of handling large online classes is only going to be higher.

8.4 Transcript-based annotations for future use

Live streaming videos and related materials are usually saved for future use. For example, it is common for learners who attended a live stream to review the content later for exam preparation. Uploading the recorded video and supplementary materials to public platforms is also adopted by many educators, aiming to distribute knowledge and mitigate education inequity. In these cases, the live streaming video becomes a pre-recorded video, and the viewers may suffer from the lack of the instructor's and peers' support.

The collected annotations made by the learners of EduLive could be leveraged as learning support for subsequent viewers of the recorded videos. For example, previous work has shown that the visualization bars reflecting areas of interest may navigate subsequent learners in learning [18]. With the annotations on top of transcript segments, which can be mapped to corresponding timestamps of the video, video interface widgets could be augmented by 2D non-linear timelines that guide learners' attention [34]. Likewise, with the context-specific highlights and questions collected using EduLive, later learners can be instructed with an additional context when learning with the videos and reading the annotations.

From another perspective, the learner's annotations are also meaningful for the instructors after the live streams. As hinted by I-SC, visualization bars associated with the transcripts intuitively revealed which parts were perceived to be interesting and which parts were perceived to be confusing, which is valuable information for future course design and lecture preparation.

From these perspectives, transcript-based annotations collected in educational live streams open up a unique design space for assisting broader audiences beyond the participants of the live streams.

8.5 Limitations and Future Work

Due to the challenging nature of hosting live streaming sessions that require many participants to join at the same time for the system evaluation, we conducted a relatively small-scale data collection. Although this approach may bring some limitations (e.g., failing to reveal how instructor/learners dynamics could be affected when there are hundreds or thousands of participants), this evaluation method suffices for us to capture instructors' and learners' key reactions and concerns to the proof-of-concept systems and the effects of involving real-time transcripts and transcript-based annotations in educational live streams. Future work can deploy EduLive in a field study to explore how instructors and learners from larger classes or different styles of classes (e.g., workshops) use it.

Another limitation comes from the background of the recruited participants in our study. Although we tried to only recruit participants who were genuinely interested in the topics of the live streams to ensure that their motivations for attending were similar to real learners in online learning scenarios, we understood that they were individuals from the online recruiting platform who did not have personal ties to other learners and the instructors, thus their interactions with each other may not be as natural as interactions in real class settings in which instructors and learners have personal connections with each other. For example, in the four classes we arranged, most interactions were between instructors and learners and we saw few learner-learner interactions. Future work can study how instructors and learners from real classes (i.e., with personal ties) use EduLive to support their interactions.

For user studies conducted with crowd workers, implementing effective attention check mechanisms is essential for collecting valid data. One notable strategy we employed was having each participant ask at least two questions. This approach not only helps hold participants' attention during live streams but also guarantees sufficient annotations for EduLive, making the aggregation process meaningful. Despite the potential limitation of adding stress on participants and possibly affecting their motivation to ask questions, we believe its impact on the evaluation's validity is minimal, given that the same attention check is uniformly applied across all conditions, and instructors are informed of this requirement. A field study with EduLive could also resolve this limitation brought by the nature of the experiment.

9 Conclusion

In this paper, we explored the opportunity of re-creating interaction cues in online educational live streams by employing real-time transcripts and transcript-based annotations. We proposed EduLive, a system that aggregates learners' annotations to assist live stream instructors in tracking learners' engagement and confusion and scaffold instructor-learners interaction. With four simulated classes, we evaluated EduLive along with two baseline systems. We found that instructors benefited from the transcript-provided context when addressing learners' questions and perceived the pausing mechanism of EduLive to be helpful in maintaining focus on the main teaching content. From the learners' perspectives, they appreciated that EduLive helped them formulate high-quality questions and follow along with the live streams effectively. We confirmed the value of employing transcript

and transcript-based annotations in educational live streaming and presented our reflections on how future designs can make better live stream supporting systems in educational settings.

References

[1] 2023. *Azure samples: cognitive services speech SDK.* https://github.com/Azure-Samples/cognitive-services-speech-sdk

[2] 2023. *OBS Studio.* https://obsproject.com/

[3] Karan Ahuja, Dohyun Kim, Franceska Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan, and Yuvraj Agarwal. 2019. EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–26.

[4] Karan Ahuja, Deval Shah, Sujeath Pareddy, Franceska Xhakaj, Amy Ogan, Yuvraj Agarwal, and Chris Harrison. 2021. Classroom digital twins with instrumentation-free gaze tracking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[6] Ciro Cattuto, Alain Barrat, Andrea Baldassarri, Gregory Schehr, and Vittorio Loreto. 2009. Collective dynamics of social annotation. *Proceedings of the National Academy of Sciences* 106, 26 (2009), 10511–10515.

[7] Di Chen, Dustin Freeman, and Ravin Balakrishnan. 2019. Integrating multimedia tools to enrich interactions in live streaming for language learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.

[8] Milton Chen. 2003. Visualizing the pulse of a classroom. In *Proceedings of the eleventh ACM international conference on Multimedia*. 555–561.

[9] Wenzhi Chen and Manlai You. 2007. The differences between the influences of synchronous and asynchronous modes on collaborative learning project of industrial design. In *International Conference on Online Communities and Social Computing*. Springer, 275–283.

[10] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. " I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.

[11] Yan Chen, Walter S Lasecki, and Tao Dong. 2021. Towards supporting programming education at scale via live streaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–19.

[12] Zhilong Chen, Hancheng Cao, Yuting Deng, Xuan Gao, Jinghua Piao, Fengli Xu, Yu Zhang, and Yong Li. 2021. Learning from home: A mixed-methods analysis of live streaming based remote education experience in Chinese colleges during the COVID-19 pandemic. In *Proceedings of the 2021 CHI Conference on human factors in computing systems*. 1–16.

[13] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, Ari Shapiro, and Stefan Scherer. 2015. Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1143–1154.

[14] Paul Denny, John Hamer, Andrew Luxton-Reilly, and Helen Purchase. 2008. PeerWise: students sharing their multiple choice questions. In *Proceedings of the fourth international workshop on computing education research*. 51–58.

[15] Danilo Dessì, Gianni Fenu, Mirko Marras, and Diego Reforgiato Recupero. 2019. Bridging learning analytics and cognitive computing for big data classification in micro-learning video collections. *Computers in Human Behavior* 92 (2019), 468–477.

[16] Teresa Duncan, Paul Pintrich, David Smith, and Wilbert Mckeachie. 2015. Motivated strategies for learning questionnaire (MSLQ) manual. *National Center for Research to Improve Postsecondary Teaching and Learning. https://doi.org/10.13140/RG* 2, 2547.6968 (2015).

[17] Sabine Fabriz, Julia Mendzheritskaya, and Sebastian Stehle. 2021. Impact of synchronous and asynchronous settings of online teaching and learning in higher education on students' learning experience during COVID-19. *Frontiers in Psychology* (2021), 4544.

[18] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, Ching Liu, and Hao-Chuan Wang. 2022. Understanding the Effects of Structured Note-taking Systems for Video-based Learners in Individual and Social Learning Contexts. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–21.

[19] Jingchao Fang, Yanhao Wang, Chi-Lan Yang, and Hao-Chuan Wang. 2021. NoteCoStruct: Powering Online Learners with Socially Scaffolded Note Taking and Sharing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–5.

[20] C Ailie Fraser, Joy O Kim, Hijung Valentina Shin, Joel Brandt, and Mira Dontcheva. 2020. Temporal segmentation of creative live streams. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[21] 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. 1555–1564.

[22] Elena L Glassman, Aaron Lin, Carrie J Cai, and Robert C Miller. 2016. Learnersourcing personalized hints. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1626–1636.

[23] Mitchell Gordon and Philip J Guo. 2015. Codepourri: Creating visual coding tutorials using a volunteer crowd of learners. In *2015 IEEE symposium on visual languages and human-centric computing (VL/HCC)*. IEEE, 13–21.

[24] Maja Grgurović and Volker Hegelheimer. 2007. Help options and multimedia listening: Students' use of subtitles and the transcript. *Language learning & technology* 11, 1 (2007), 45–66.

[25] William A Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1315–1324.

[26] William A Hamilton, Nic Lupfer, Nicolas Botello, Tyler Tesch, Alex Stacy, Jeremy Merrill, Blake Williford, Frank R Bentley, and Andruid Kerne. 2018. Collaborative live media curation: Shared context for participation in online learning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[27] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 Chi conference on human factors in computing systems*. 5114–5119.

[28] Thomas T Hills. 2015. Crowdsourcing content creation in the classroom. *Journal of Computing in Higher Education* 27, 1 (2015), 47–67.

[29] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. 2017. Intelligent Tutors as Teachers' Aides: Exploring Teacher Needs for Real-Time Analytics in Blended Classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (Vancouver, British Columbia, Canada) *(LAK '17)*. Association for Computing Machinery, New York, NY, USA, 257–266. https://doi.org/10.1145/3027385.3027451

[30] Stefan Hrastinski. 2008. Asynchronous and synchronous e-learning. *Educause quarterly* 31, 4 (2008), 51–55.

[31] Min Hu and Hao Li. 2017. Student engagement in online learning: A review. In *2017 International Symposium on Educational Technology (ISET)*. IEEE, 39–43.

[32] Hassan Khosravi, Kirsty Kitto, and Joseph Jay Williams. 2019. Ripple: a crowdsourced adaptive platform for recommendation of learning activities. *arXiv preprint arXiv:1910.05522* (2019).

[33] Juho Kim et al. 2015. *Learnersourcing: improving learning with collective learner activity*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[34] Juho Kim, Philip J Guo, Carrie J Cai, Shang-Wen 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*. 563–572.

[35] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2009. What's in Wikipedia?: mapping topics and conflict using socially annotated category structure. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, 2009*, Dan R. Olsen Jr., Richard B. Arthur, Ken Hinckley, Meredith Ringel Morris, Scott E. Hudson, and Saul Greenberg (Eds.). ACM, New York, NY, USA, 1509–1512.

[36] Nataliya Kosmyna and Pattie Maes. 2019. Attentivu: An EEG-based closed-loop biofeedback system for real-time monitoring and improvement of engagement for personalized learning. *Sensors* 19, 23 (2019), 5200.

[37] Charlotte Kramer, Johannes König, Sarah Strauss, and Kai Kaspar. 2020. Classroom videos or transcripts? A quasi-experimental study to assess the effects of media-based learning on pre-service teachers' situation-specific skills of classroom management. *International Journal of Educational Research* 103 (2020), 101624.

[38] Chinmay Kulkarni and Ed Chi. 2013. All the news that's fit to read: a study of social annotations for news reading. In *2013 ACM SIGCHI Conference on Human Factors in Computing Systems, CHI '13, Paris, France, April 27 - May 2, 2013*, Wendy E. Mackay, Stephen A. Brewster, and Susanne Bødker (Eds.). Association for Computing Machinery, New York, NY, USA, 2407–2416.

[39] Raja S Kushalnagar, Walter S Lasecki, and Jeffrey P Bigham. 2013. Captions versus transcripts for online video content. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*. 1–4.

[40] Walter S Lasecki, Juho Kim, Nick Rafter, Onkur Sen, Jeffrey P Bigham, and Michael S Bernstein. 2015. Apparition: Crowdsourced user interfaces that come to life as you sketch them. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1925–1934.

[41] Hsien-Chin Liou. 1997. Research of on-line help as learner strategies for multimedia CALL evaluation. *Calico Journal* (1997), 81–96.

[42] Ching Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. Notestruct: Scaffolding note-taking while learning from online videos. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.

[43] Zhicong Lu, Seongkook Heo, and Daniel J Wigdor. 2018. Streamwiki: Enabling viewers of knowledge sharing live streams to collaboratively generate archival documentation for effective in-stream and post hoc learning. *Proceedings*

of the ACM on Human-Computer Interaction 2, CSCW (2018), 1–26.

[44] Zhicong Lu, Rubaiat Habib Kazi, Li-Yi Wei, Mira Dontcheva, and Karrie Karahalios. 2021. StreamSketch: Exploring Multi-Modal Interactions in Creative Live Streams. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[45] Shuai Ma, Taichang Zhou, Fei Nie, and Xiaojuan Ma. 2022. Glancee: An Adaptable System for Instructors to Grasp Student Learning Status in Synchronous Online Classes. In *CHI Conference on Human Factors in Computing Systems*. 1–25.

[46] Albert Malkin, Ruth Anne Rehfeldt, and Ashley M Shayter. 2018. An investigation of the efficacy of asynchronous discussion on students' performance in an online research method course. *Behavior analysis in practice* 11, 3 (2018), 274–278.

[47] Steven Moore, Huy A Nguyen, and John Stamper. 2020. Evaluating Crowdsourcing and Topic Modeling in Generating Knowledge Components from Explanations. In *International Conference on Artificial Intelligence in Education*. Springer, 398–410.

[48] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. Affectivespotlight: Facilitating the communication of affective responses from audience members during online presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[49] Mei-Hua Pan, Naomi Yamashita, and Hao-Chuan Wang. 2017. Task rebalancing: Improving multilingual communication with native speakers-generated highlights on automated transcripts. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 310–321.

[50] Amy Pavel, Colorado Reed, Björn Hartmann, and Maneesh Agrawala. 2014. Video digests: a browsable, skimmable format for informational lecture videos.. In *UIST*, Vol. 10. Citeseer, 2642918–2647400.

[51] Amy T Peterson, Patrick N Beymer, and Ralph T Putnam. 2018. Synchronous and asynchronous discussions: Effects on cooperation, belonging, and affect. *Online Learning* 22, 4 (2018), 7–25.

[52] Sandra Poeschl and Nicola Doering. 2015. Measuring co-presence and social presence in virtual environments–psychometric construction of a german scale for a fear of public speaking scenario. *Annual Review of Cybertherapy and Telemedicine 2015* (2015), 58–63.

[53] Bethany Rittle-Johnson. 2006. Promoting transfer: Effects of self-explanation and direct instruction. *Child development* 77, 1 (2006), 1–15.

[54] Amy Shannon, Jessica Hammer, Hassler Thurston, Natalie Diehl, and Steven Dow. 2016. PeerPresents: A Web-Based System for In-Class Peer Feedback during Student Presentations. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. Association for Computing Machinery, New York, NY, USA, 447–458. https://doi.org/10.1145/2901790.2901816

[55] Lijie Shao, Fuwei Zhang, Ruomei Wang, Fan Zhou, and Shujin Lin. 2018. An Efficient Expansion Word Extraction Algorithm for Educational Video. In *2018 7th International Conference on Digital Home (ICDH)*. IEEE, 131–136.

[56] Anjali Singh, Christopher Brooks, and Shayan Doroudi. 2022. Learnersourcing in Theory and Practice: Synthesizing the Literature and Charting the Future. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 234–245.

[57] Hendrik Steinbeck, Ralf Teusner, and Christoph Meinel. 2021. Teaching the Masses on Twitch: An Initial Exploration of Educational Live-Streaming. In *Proceedings of the Eighth ACM Conference on Learning@ Scale*. 275–278.

[58] Wei Sun, Yunzhi Li, Feng Tian, Xiangmin Fan, and Hongan Wang. 2019. How Presenters Perceive and React to Audience Flow Prediction In-situ: An Explorative Study of Live Online Lectures. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.

[59] Ilaria Torre, Ilenia Galluccio, and Mauro Coccoli. 2022. Video augmentation to support video-based learning. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. 1–5.

[60] Claire E Weinstein and Richard E Mayer. 1983. The teaching of learning strategies.. In *Innovation abstracts*, Vol. 5. ERIC, n32.

[61] Sarah Weir, Juho Kim, Krzysztof Z Gajos, and Robert C Miller. 2015. Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 405–416.

[62] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z Gajos, Walter S Lasecki, and Neil Heffernan. 2016. Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. 379–388.

[63] Man Wu and Qin Gao. 2020. Using live video streaming in online tutoring: Exploring factors affecting social interaction. *International Journal of Human–Computer Interaction* 36, 10 (2020), 964–977.

[64] Françeska Xhakaj, Vincent Aleven, and Bruce M McLaren. 2016. How teachers use data to help students learn: Contextual inquiry for the design of a dashboard. In *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, Lyon, France, September 13-16, 2016, Proceedings 11*. Springer, 340–354.

[65] Saelyne Yang, Changyoon Lee, Hijung Valentina Shin, and Juho Kim. 2020. Snapstream: Snapshot-based interaction in live streaming for visual art. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[66] Saelyne Yang, Jisu Yim, Juho Kim, and Hijung Valentina Shin. 2022. CatchLive: real-time summarization of live streams with stream content and interaction data. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–20.

[67] Matin Yarmand, Jaemarie Solyst, Scott Klemmer, and Nadir Weibel. 2021. "It Feels Like I am Talking into a Void": Understanding Interaction Gaps in Synchronous Online Classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.

[68] Salu Ylirisku, Giyong Jang, and Nitin Sawhney. 2021. Re-thinking pedagogy and dis-embodied interaction for online learning and co-design. *Nordes* 1, 9 (2021).

[69] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful classroom deployment of a social document annotation system. In *Proceedings of the sigchi conference on human factors in computing systems*. 1883–1892.

A Survey instruments measuring learning experience

In Table 5, we list survey questions (presented on a 7-point Likert scale) used in our study.

In addition, we also asked open-ended questions to collect learners' detailed feedback. These questions include *"Do you have any other feedback about your learning experience that you would like to share with us?"*, *"Did you [use the transcript/make use of the visualizations bars around the transcript/make use of summarization box when the streamer was pausing]? If yes, how did you use it, and how did it positively/negatively affect your experience? If not, why?"*.

Received July 2023; revised January 2024; accepted March 2024

Themes	***Survey questions***
Co-presence with other learners	*"I feel like other learners are with me in the live streaming class."* *"I was aware of other learners' presence in the live streaming class."* *"I feel like other learners know that I'm in the live streaming class."*
Easiness of participating and asking questions	*"I can easily participate in this live streaming class."* *"Participating in the class takes a lot of time and effort."* *"Asking questions using the interface is easy."* *"I'm not sure if my questions are clear and understandable to others."*
Easiness of sharing and expressing to peers	*"It is easy for me to share my opinions with peers."* *"I am aware of what my peers shared with me."* *"I have the feeling that my peers and I can express our thoughts to each other easily."*
Understanding of questions and answers in the live stream	*"When I look into my peers' questions, I understand what they are asking about."* *"When I look into my peers' questions, I don't know what part of the class they were referring to."* *"When the instructor is answering questions, I know what he/she is talking about."* *"I know the context of the questions when the instructor is answering them."*
Reviewability of the live stream	*During the live streaming class, I can review the course content from several minutes ago.* *"I can easily trace back to the content of a certain part of the live streaming class."*
Sense of agency	*"I feel that my view and opinions are taken into account in this class."* *"I can influence the class."* *"I feel that my viewpoints were listened to."* *"I feel it is impossible for me to influence the class."*
Self-efficacy of learning	*"I'm certain I can master the knowledge being taught in the live stream."* *"I'm confident I can understand the basic concepts taught in the live stream. "* *"I'm confident I can understand the most complex material presented by the instructor in the live stream."*

Table 5. Survey instruments