XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences

Michael Nebeling[1], Shwetha Rajaram[1], Liwei Wu[1], Yifei Cheng[1,2], Jaylin Herskovitz[1]
[1] University of Michigan [2] Swarthmore College
[nebeling,shwethar,wuliwei,yifcheng,jayhersk]@umich.edu

ABSTRACT

There is increased interest in using virtual reality in education, but it often remains an isolated experience that is difficult to integrate into current instructional experiences. In this work, we adapt virtual production techniques from filmmaking to enable mixed reality capture of instructors so that they appear to be standing directly in the virtual scene. We also capitalize on the growing popularity of live streaming software for video conferencing and live production. With XRStudio, we develop a pipeline for giving lectures in VR, enabling live compositing using a variety of presets and real-time output to traditional video and more immersive formats. We present interviews with media designers experienced in film and MOOC production that informed our design. Through walkthrough demonstrations of XRStudio with instructors experienced with VR, we learn how it could be used in a variety of domains. In end-to-end evaluations with students, we analyze and compare differences of traditional video vs. more immersive lectures with XRStudio.

CCS CONCEPTS

• **Human-centered computing → Interface design prototyping**; **Mixed / augmented reality**.

KEYWORDS

virtual production; mixed reality capture; immersive instruction.

ACM Reference Format:
Michael Nebeling[1], Shwetha Rajaram[1], Liwei Wu[1], Yifei Cheng[1,2], Jaylin Herskovitz[1]. 2021. XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3411764.3445323

1 INTRODUCTION

While there is a growing interest in using virtual reality to support teaching and instruction [25, 26, 33], there is relatively limited tool support, let alone an easy and intuitive workflow, to facilitate this. Often only instructors with significant technical skills and those who are well versed in tools like Unity can accomplish this task on their own [1, 31, 32]. Most will have to work with a team of designers and developers to create VR solutions tailored to their teaching needs. The goal of this work is not to transform all instructional tasks to VR or demonstrate the value of teaching in VR—this is the focus of other research [42]. Rather, we want to enable instructors who wish to teach in VR and students who lack access to VR. For most students, there is both a significant learning curve to use VR technologies and also the need to imagine how the person in VR actually perceives the content if they themselves cannot directly participate in the virtual experience [14, 29, 36].

We think of this work as systems research on how to make it easier for both instructors and students to have more immersive educational experiences. We capitalize on the growing popularity of live streaming software such as *OBS Studio*, *Streamlabs OBS*, or *XSplit*, which integrate with established online video platforms such as *YouTube* and *Twitch*. The key idea behind our work is to adapt virtual production workflows and tools to create new solutions for immersive teaching and learning. Virtual production is a new technique increasingly used in filmmaking to create virtual sets and special effects in a live composition rather than being added in post-production. Compared to using green screen backgrounds, the technique is characterized by mixing physical and virtual filmsets made from 3D digital content, making the virtual objects directly visible to the actors, and recording the mixed reality live composition in real-time. Popularized through recent productions such as the *Lion King* remake, *First Man*, or *The Mandalorian*, this new immersive filming technique enables everyone on set to get an accurate preview of the final viewing experience and to have direct control over it even during production. However, this typically requires extensive instrumentation of the studio stage and the actors with advanced display and tracking technologies, which is not feasible in most educational settings.

In this paper, we present *XRStudio*, a system we designed to investigate how to adapt virtual production techniques via mixed reality capture to enable immersive instructional experiences without expensive technical setups or production teams. As a system, XRStudio shares techniques with new emerging tools such as *LIV* and *MixCast*, which support mixed reality capture and live compositing of a user playing VR games such as *Beat Saber* or using immersive sketching tools such as *Tilt Brush*. Our system goes beyond capture, allowing us to study the entire workflow from creating immersive content as an instructor to accessing the content as a student. XRStudio structures the workflow into four steps: *(1)* calibration and integration of multiple physical and virtual camera angles, *(2)* mixed reality capture with background removal to segment the instructor from live video feeds and facilitate depth perception in relation to the virtual content, *(3)* import/export of instructional materials such as 2D slides and 3D models through a flexible whiteboard component, and *(4)* record-and-replay of immersive experiences in a variety of lecture formats, ranging from

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445323

2D video and 3D scenes, to immersive VR and AR sessions. Having implemented the entire workflow, we use XRStudio to study the new affordances with live video streaming of mixed reality capture from the instructor to students, and analyze the learning experience and production quality comparing traditional video to the more immersive lecture formats enabled by XRStudio.

This paper establishes principles for virtual production and live streaming of immersive instructional experiences. We present the systems research behind XRStudio, detailing the pipeline and implementation of the technical components part of the larger production workflow. We report on the design and technical challenges and how we addressed each when creating our prototype system. Then, using XRStudio, we gather insights from user studies with three main stakeholders: *(1)* **media designers** responsible for the production of lecture recordings in both residential and online learning who shared their current workflows and challenges in a focus group, *(2)* **instructors** in different domains who participated in interviews to discuss their strategies of incorporating VR into lecture and classroom activities, and who brainstormed with us based on a system walkthrough of XRStudio how they might use versions of it in their teaching, and *(3)* **students** in an interaction design course focused on AR/VR technologies who provided feedback on the end-to-end user experience. Based on these studies, we see great potential for XRStudio to increase student engagement in virtual and remote teaching and learning, which is particularly relevant during the COVID-19 pandemic. Finally, we discuss issues such as selecting good lecture topics for immersive instructional experiences, balancing instructor effort, and shaping best practices around systems like XRStudio, as well as future studies to address limitations.

2 RELATED WORK

Our work on XRStudio bridges and adds to three streams of research: *VR in education*—attempting to define the use and role of VR for teaching and instruction, *sharing VR experiences*—developing techniques to show to non-HMD users what HMD users are experiencing, and *immersive authoring and virtual production*—investigating tools to create and capture virtual content while being in VR.

2.1 VR in Education

There is a large corpus of works investigating the unique affordances that VR provides for education [2, 3, 12, 18, 20, 30, 34, 39, 40, 42]. Two commonly demonstrated benefits of VR in education are: *(1)* enabling students to explore environments and experiences that would either be difficult or impossible to access [7, 20, 29, 34] (e.g., surgery training environments [23]), and *(2)* supporting understanding of spatial content by situating students in the same context, enabling them to interact with the content in a more direct manner [24, 25, 33, 41].

While there are many compelling applications of VR, developing VR learning experiences remains a challenge. Currently, commercial VR authoring tools such as Unity still require users to know programming. Existing tools for creating VR experiences without the need for programming are primarily targeted at designers and not optimized for educational purposes [31, 32]. With XRStudio, we aim to lower the barrier to entry for instructors to teach with VR through providing a virtual classroom environment that is configurable without code and a set of flexible teaching tools, including virtual whiteboards with 2D/3D sketching tools.

2.2 Sharing VR Experiences

The ability to share VR experiences with non-HMD users (spectators) is important, particularly in educational scenarios, to ensure students are not denied learning opportunities due to lack of access to VR equipment. However, as most VR systems primarily focus on the HMD user experience, VR remains mainly a closed experience among those experiencing the VR in first person via a headset [14, 15]. Prior work explored different approaches to enabling people without VR equipment to participate in immersive experiences. These include mirroring the first-person view of the HMD user [29], providing access to the VR user's virtual world via a PC interface [11, 21], projecting the immersive environment around the HMD user [14, 22], and mounting peripheral devices onto the HMD to allow for interactions between HMD and non-HMD users [4, 15].

Most of these systems assume that HMD and non-HMD users are co-located in the same physical environment, but this is not always realistic in educational contexts. If approaches are limited this way, there are major issues of accessibility and equity posed to those who do not have the physical ability, experience nausea, or lack the financial means to purchase personal HMDs. We designed XRStudio with both co-located and remote teaching scenarios in mind. Similar to [11] and [21], we support experiencing immersive content through multiple modalities, including 3D desktop, VR HMDs, and mobile AR. With XRStudio, we study a new way of sharing VR experiences via virtual production techniques with presets for live compositing and streaming. We hope to improve upon the shared-screen approach to VR spectator views, by providing a wider range of options for sharing immersive content, from first-person and third-person VR, to mixed reality capture.

2.3 Immersive Authoring & Virtual Production

This paper builds upon virtual production techniques from prior research and industry examples, adapting them to an instructional context. Recent film productions increasingly employ virtual production techniques, but these require extensive instrumentation and are therefore infeasible to replicate in most educational contexts. Emerging tools such as *LIV* and *MixCast*[1] are gradually enabling virtual production techniques with more lightweight setups. XRStudio takes the capture techniques in LIV and MixCast further to support real-time compositing of lecture recordings with virtual elements. We build on recent automatic person segmentation methods to reduce the physical setup required for XRStudio, extracting the instructor from their live video feed without a green screen.

While the term *"virtual production"* is rarely used in HCI research, several prior works implement related techniques. KinÊtre [6] enabled users to use their bodies to animate scanned virtual objects and 3D Puppetry [19] operated 3D models on a virtual set in real-time using physical props. XRDirector [31] expanded these puppeteering techniques by mapping user positions to virtual elements like 3D characters, lights, and cameras, and film interactions in AR or VR. LACES [13] enabled users to edit videos live

[1] https://liv.tv; https://mixcast.me

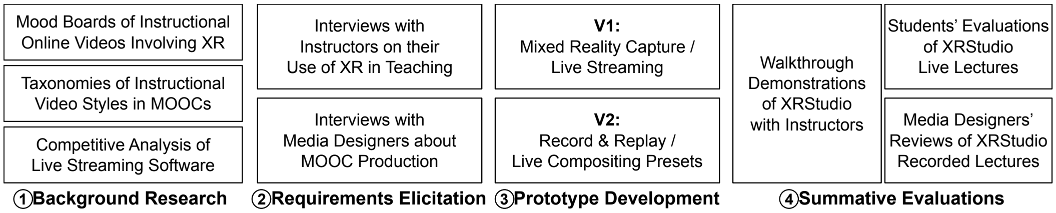

Figure 1: Our research process can be divided into four stages: *(1)* background research to scope our research and guide system design, *(2)* requirements elicitation with instructors and media designers as key stakeholders, *(3)* prototype development to create the system pipeline and support testing, and *(4)* summative evaluations of XRStudio walkthrough demonstrations with instructors, and lectures with students and media designers.

with different composition tools to reduce post-processing time. Saquib et al.'s augmented video performance system [38] enabled live composition of recorded video and gesture-triggered graphics. SAVE [43], Loki [25], and TutoriVR [26] enabled live sharing of VR experiences while MAVRC [8] and TransceiVR [27] studied asynchronous and asymmetric collaboration between HMD and non-HMD users. All these tools share similarities with XRStudio but were limited to supporting only one camera perspective and capture modality at a time. XRStudio combines virtual production and live streaming, enabling an instructor to make use of multiple camera angles and capture modalities via presets such that students can obtain fully virtual as well as mixed reality captured views of the lecture material, which broadens access to VR content.

3 BACKGROUND RESEARCH

Figure 1 shows our research process. This section details the background research we conducted in three steps: *(1a)* creating mood boards of instructional videos involving some form of XR, *(1b)* consolidating existing taxonomies of instructional video styles in MOOCs, and *(1c)* competitive analysis of existing streaming platforms. Later sections will detail subsequent steps in our research.

3.1 Mood Boards of Instructional Recordings

Early on in the research, all members of our team independently created mood boards around innovative uses of XR technologies in teaching and instruction. The methods for seeking videos varied; some identified relevant sources by searching the terms {AR, VR, XR} ∪ {Education, Classroom} on YouTube, others chose to browse through video galleries created around topics that could benefit from 3D visualization or involved an instructor somehow composed with virtual content. In this phase of the research, we wanted to be more explorative than systematic and benefited from this variety. When we came together to share and analyze our mood boards, we had found a total of 28 inspiring examples.

We categorized the examples into four groups: **fully virtual** (e.g., sharing first-person VR, or representing the instructor as a virtual avatar in a 3D world), **virtual with spatially placed video streams** (e.g., composing a real, life-sized instructor with a virtual world, or inserting a video stream of the instructor into a virtual world), **video with spatially placed virtual content** (e.g., overlaying virtual content on a video stream of the real world, or embedding virtual content in the spatially appropriate location within the real world), and **virtual separate from video** (e.g., superimposing the instructor video stream with first-person VR, or representing the instructor video feed and first-person VR in a split-screen setup).

From the examples, we also derived a preliminary set of features to implement in XRStudio. First, we were inspired to create presets for common video production styles. Second, we wanted to enable live compositing of an instructor's video feed with the virtual content via mixed reality capture. Finally, we identified a common set of tools we wanted to implement in XRStudio for content authoring and annotation in VR leading to our flexible whiteboard concept supporting slides and 2D sketches as well as 3D modeling.

3.2 Taxonomies of Instructional Video Styles

We envisioned XRStudio as a versatile video production tool, capable of supporting a wide variety of instructional styles. To validate and extend our design features obtained from preliminary mood boards, we analyzed existing massive open online courses (MOOCs) as potential examples of high-quality instructional video productions. We studied existing taxonomies of MOOC video composition styles, which has been an active area of research over the past few years, and identified six taxonomies for instructional video styles from [5, 9, 16, 17, 35, 37]. Together, these taxonomies contained a total of 73 instructional video styles, which we categorized into seven groups: *(1)* **talking head** (i.e., upper-body crop of the instructor lecturing into the camera) [9, 16, 17, 37], *(2)* **live lecture recording** (i.e., lecture recording from the perspective of a student) [5, 9, 16, 17, 37], *(3)* **interview** (i.e., recording of a dialogue) [9, 17, 37], *(4)* **slides** (i.e., presentation slides with narration) [5, 16, 35, 37], *(5)* **whiteboard** (i.e., stream of an instructor freehand drawing on a surface) [9, 16, 17, 35, 37], *(6)* **screen-cast** (i.e., stream of an instructor's screen) [9, 16, 17, 35, 37], *(7)* **miscellaneous** (e.g., custom animation, demonstration recordings, location shoots) [17, 35, 37]. We implemented versions of these video production styles in XRStudio, with the exception of those in the miscellaneous category since they were much less common.

3.3 Analysis of Live Streaming Software

To identify a list of key features that a live streaming solution in XRStudio would require, we compiled a list of nine streaming software platforms. We selected four of the most popular ones for a competitive analysis: *OBS Studio*, *Streamlabs OBS*, *XSplit*, and *Lightstream*. Common to all applications was support for streaming directly to platforms like YouTube, Twitch, Facebook Live, and Periscope. These systems also commonly enable recording while streaming. We determined that this would likewise be a desirable feature for XRStudio. Through a notion of *presets*, streaming platforms provide users with a lot of flexibility in setting up their stream using a variety of scene composition tools, enabling multiple media sources to be freely combined and arranged. Users are also given control over the stream via a set of hotkeys during live production, e.g., for showing or hiding elements in the scene composition or switching scenes entirely. For XRStudio, we thought that instructors would benefit from having a comparable amount of freedom, and set as a design requirement that the instructional experience be easily configurable, both in terms of the VR experience as well as the live production, via scene templates and presets. Another frequently-used feature of streaming software was chat. In videoconferencing software like Zoom, this extends from text to audio/video chat, basic versions of which we implemented in XRStudio.

4 REQUIREMENTS ELICITATION

In the formative stages of XRStudio, we interviewed media designers and instructors to elicit requirements from key stakeholders.

4.1 Production Requirements

To learn about video production requirements, we conducted interviews with a team of two media design experts and their technology lead who had several years of experience in educational and professional media production. We recruited the participants from an on-campus unit focused on innovating in residential and online learning experiences. Our goal was to get an overview of their MOOC production workflow and identify opportunities to integrate immersive technologies with their current process and tools.

We structured the interview into four blocks: *(1)* overview of the participants' prior production experience and skill sets; *(2)* walkthrough of their production workflows including the technologies they utilized and production roles involved at each step of the process; *(3)* discussion on the challenges with their current workflows, in particular, transitioning to remote due to COVID-19; and *(4)* brainstorming session on how immersive technology could be used to improve the production process and learning experience for students. We aggregated the responses and extracted three requirements, which informed the design of our XRStudio system.

R1: High-quality video without a production team. The media designers shared that most learning experiences they developed involved production teams of highly specialized media designers with diverse video editing skills. They identified post-production as a crucial step in the process, typically demanding five hours of video editing for one hour of raw video. However, this level of effort and specialized skills would be infeasible for instructors who also need to invest time into developing lecture content. As such, we aimed to enable instructors to produce high-quality video on their own using XRStudio, by providing a *default scene and pre-calibrated cameras* angled towards the instructor and lecture content.

R2: Automating the production process. The participants expressed that automating hardware and software configurations was key for production efficiency, considering that they typically work on 4–6 projects in parallel. The tech lead stated that 90% of their work was produced with standard video templates and that they use three standard room setups with different lighting and equipment to support common MOOC video composition styles. We adopted a similar approach with XRStudio to streamline the production process, by providing *live production presets* for live compositing of the instructor and lecture content in various styles.

R3: "No-technology" solution for instructors. The media designers agreed that it is best to involve instructors directly in the production process, since they are the experts on the course content and may have particular visions for the production. However, many instructors struggle with handling the technological infrastructure on their own. The tech lead expressed that instructors "should not need to know anything at all about the technology." With XRStudio, we simplified the hardware and software setup, by providing *mixed reality capture* without the need for instrumentation of the physical environment (e.g., green screen) or the user (e.g., fiducial) based on automatic person segmentation and a three-point calibration technique to map the instructor's webcam to a virtual camera.

4.2 Instructional Requirements

To inform the design of XRStudio from the instructor perspective, we conducted five semi-structured interviews with instructors who have used XR in their teaching. We recruited via an on-campus working group of instructors engaged in teaching and learning with XR. Our goal was to learn about their strategies to incorporate XR into learning activities and any challenges they faced. Each instructor worked in a different domain: nursing, film and theater, architecture, landscape architecture and sustainability, and learning experience design. Their course sizes ranged from six students for design studio courses to 80–100 students for larger lectures.

Each interview consisted of two blocks: *(1)* a discussion on the instructor's use of XR for teaching, where we focused on a specific classroom activity and the content creation process, and *(2)* a reflection on challenges encountered with integrating XR into the classroom and a discussion on what might be the "perfect tool" which could help to mitigate these challenges.

Instructors' use of XR was two-fold: *(i)* simulating interactivity and complex concepts, and *(ii)* creating and visualizing immersive content. First, simulation-based activities included multi-user training applications for medical procedures which utilize both VR and AR, a VR architecture "textbook" which teaches material properties and 2D sketching of 3D structures, and a VR literary experience used for exploring examples of educational games. As examples of the second category, one instructor led a course on VR production, teaching students in film and media to create 360 VR experiences. Students in a landscape architecture course used VR to visualize their 3D environment models to gain an understanding of perspective and scale. We extracted themes from the interviews and identified three additional requirements for XRStudio.

R4: Autonomy in VR content creation. Some of the instructors who utilized simulation-based activities worked with specialized XR development teams at the university to develop the experiences, due to their own time limitations and lack of a programming background. While satisfied with these tailored solutions, they cited challenges communicating their specific vision for the experience to developers who often lacked critical domain knowledge, as well as encountering delays in development. Two opted to use commercial XR experiences but they still felt a need for more control over the design. To many instructors, a "perfect tool" for XR instruction would be a "no-code development environment," allowing them to import 3D objects into the scene and adding interactivity via visual authoring. To provide instructors with more autonomy in VR content creation, we implemented a flexible set of *sketching and manipulation tools* in XRStudio, allowing instructors to visually author the VR environment and lecture content without writing any code. XRStudio's *record-and-replay system* supports time-based annotations and changes, making it possible to sketch interactive behaviors and manipulate the scene similar to many simulations.

R5: Guided learning experience in VR. While running XR activities in the classroom, a majority of instructors described serving as a "tour guide," joining the XR experience with their students whenever possible to lead them through the activities and deliver live verbal instruction while being in VR. The nursing instructor had a facilitator interface where she could help organize participants and assign roles. The film instructor guided students through techniques for developing immersive media experiences through screen sharing during live coding sessions. The architecture instructor even integrated pre-recorded voice-overs into the VR experience, to guide students when they completed the simulation exercises asynchronously. The instructors often stated a goal of giving students more independence during XR activities, providing them instructions for what to do beforehand and a framework for reflection afterwards via post-task writing assignments or critique sessions. These findings indicated a need for flexible guidance mechanisms geared towards both synchronous and asynchronous learning activities. In XRStudio, we assigned the instructor avatar a unique color to increase visibility. The *record-and-replay system* supports visual annotations and verbal instructions during capture, any portion of which can be replayed even during a new live lecture.

R6: Alternative modes of access to VR lectures. Student engagement with XR learning experiences was overwhelmingly positive. Many students were excited when "they realized that XR was an option for them." Compared to more traditional educational technology, XR still had a "wow factor" on many students due to the relative novelty of the technology. While instructors were optimistic about the role of XR for teaching, they also shared many concerns around accessibility and equity. Instructors emphasized that VR learning activities must be incorporated "with intentionality"—if it could not be accomplished "in an accessible and equitable way," they would not do it. Two instructors made the VR portions of their courses optional to accommodate students who experienced nausea. All instructors found availability of XR devices to be limited, which resulted in students having to share devices or visit one of the few VR labs on campus. To mitigate this, XRStudio offers *multiple output modalities* to reach students, including live streaming of lecture video or directly joining the 3D scene on desktop, as an alternative to joining on a VR HMD. Students can receive content live or asynchronously, as enabled by the *record-and-replay system.* XRStudio's live production presets utilize different modalities, including *mixed reality capture* to obtain a third-person perspective in addition to the instructor's first-person view in VR, as an attempt to maintain non-HMD users' sense of immersion and engagement.

5 XRSTUDIO

XRStudio enables instructors to deliver lectures in VR by providing a virtual film studio with live production facilities. It supports multiple ways to capture VR experiences including mixed reality capture, allowing students to view the instructor inside the VR experience in a composite view. Figure 2 provides a system overview, detailing the components part of our five-step pipeline for live compositing of the instructor's view while streaming to students.

5.1 Design Process

Based on the interviews with media designers and instructors, we designed XRStudio in two stages.

First, we created a system prototype to establish the pipeline for *mixed reality capture* illustrated in Figure 2. This included developing each step, from acquiring the instructor's webcam feed, person segmentation for background extraction, calibrating the mixed reality capture mapping the webcam to a virtual camera such that the VR controller models match the physical motion controllers in the real world, to live compositing of the instructor's webcam video with the VR experience. We used this version in walkthrough demonstrations of XRStudio with instructors (Section 6).

Second, we developed our *record-and-replay system* and expanded the live streaming support to include the full set of presets for the instructor to control the live compositing of themself with the scene according to their instructional activities. We used this final version of XRStudio in our end-to-end user evaluations based on live lectures with students and the corresponding lecture recordings with the media designers we had previously interviewed (Section 7).

5.2 System Walkthrough

We base our system walkthrough in this section on the scenario of an instructor giving a lecture on 2D vs. 3D design as part of an AR/VR interaction design class (Figure 3). This lecture is frequently taught by a member of the research team.

Default Scene with Pre-calibrated Cameras *(R1, R4).* First, the instructor needs to bring his lecture content into XRStudio and configure the virtual learning environment. In our example, he chooses the default scene provided by XRStudio, which uses a simple environment with a whiteboard for slides; two lights—ambient and directional, pointing center stage; and three virtual cameras, one following the instructor, one for close-ups of the slides, and one with a wider angle on the active whiteboard for any 3D objects and sketches in the scene. Our instructor embeds his slides previously exported to PNG, and loads a set of common 3D primitives. As the instructor enters VR, he can immediately start manipulating the scene content with the motion controllers. This includes controlling and annotating the slides, moving, rotating, and resizing 3D objects, and adjusting virtual camera positions. These behaviors are pre-scripted in XRStudio and can be easily toggled globally for the scene or individually for each 3D object.

Live Production Presets *(R1, R2, R3).* At this stage, the instructor can already start the lecture and enable live streaming of his webcam. XRStudio implements seven presets to support the lecture styles from our background research on MOOC videos (Section 3.2): *(1)* the webcam video feed only (**talking head**); *(2)* a wide-angle

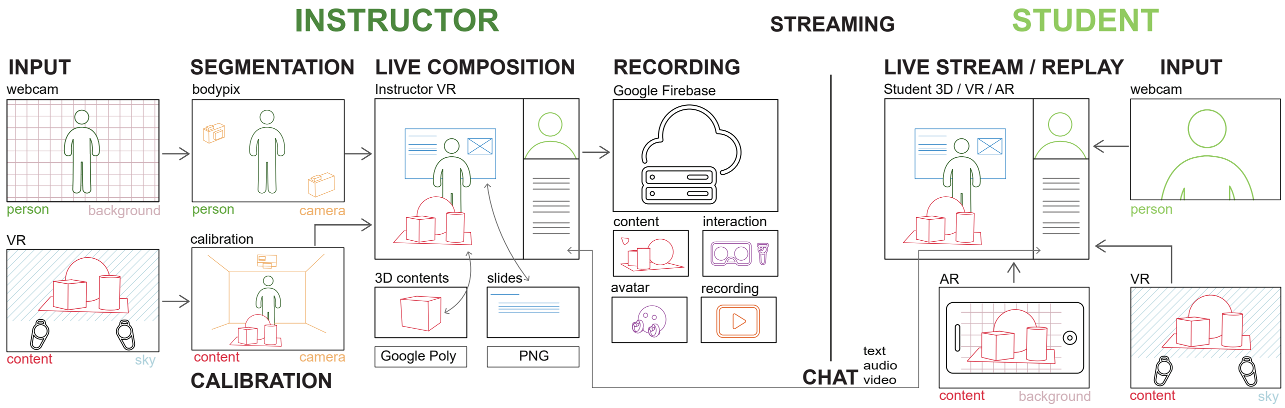

Figure 2: XRStudio comes with two clients, one for the instructor and one for students. The instructor client implements our five-steps pipeline: *1)* input—it takes the instructor's webcam video feed and the VR scene as input, *2)* segmentation—it removes the background via person segmentation, *3)* calibration—it calibrates the virtual camera to match the webcam in VR, *4)* live compositing—it composes the instructor's live video with the virtual content, and *5)* recording—it records the instructor's interactions with the lecture scene containing 3D objects and slides (PNG images). This is done in real-time at 60 FPS. At the end of a lecture, it saves the composition video and recorded data to the cloud. The student client joins the instructor's hosted lecture scene and enables the students to view a live or recorded lecture in 3D, VR, and AR. The two clients can communicate with each other through text, audio and video during a lecture.

shot on the 3D content (**live lecture recording**); *(3)* the instructor's webcam juxtaposed to live video from a student who is asking a question (**interview**); *(4)* a close-up of the slides (**slides**); *(6)* a wide-angle shot on the active 2D/3D whiteboard (**whiteboard**); *(6)* the instructor's first-person point of view (**screen-cast**); *(7)* **mixed reality capture** of the instructor with the webcam mapped to a 3D position in the scene—this requires the calibration steps detailed below. These presets render the active camera perspective in the live stream. Each preset is mapped to a number key on the keyboard (hotkey) and can also be selected via a drop-down menu in the live compositing interface, making it very easy for the instructor to switch camera angles and capture modalities for live production.

Mixed Reality Capture *(R3, R6).* To begin the lecture, the instructor sets up XRStudio's mixed reality capture via a three-point calibration technique illustrated in Figure 4. Using just a few steps, he can map the webcam to a virtual camera in the scene. During this process, he sees a virtual mirror of himself with a representation of the VR controllers to guide the calibration. Optionally, he can decide to make use of XRStudio's built-in person segmentation to extract the background from the webcam feed so that he stands out more to the students. This comes at a higher computational cost but gives more flexibility with the live production of lecture video.

Sketching & Manipulation Tools *(R4, R5).* When the instructor normally teaches the class, he often annotates slides and sketches 3D concepts on the whiteboard. XRStudio offers a variety of sketching tools to facilitate this in VR. Via a menu, the instructor can select different brushes for slide annotation and mid-air sketches. XRStudio implements three types of whiteboards (Figure 5): *2D whiteboards* hold slides, annotations, and 2D sketches; *2D/3D lightboards* allow instructors to draw on a fixed plane or in 3D space without restrictions; *3D artboards* can be used to organize 3D sketches into containers that can be manipulated in 3D physical space. These whiteboards flexibly supports free hand sketching in both 2D, as in presentation tools like PowerPoint, and 3D, as in immersive authoring tools like TiltBrush. XRStudio also implements undo/redo and supports exporting the scene content in the common glTF format.

Record-and-Replay System *(R4, R5, R6).* To enable the instructor to pre-record a lecture or capture an entire class, XRStudio implements a record-and-replay system for the instructor's webcam and live composited video, as well as the movements and actions of the instructor and students, including the advancing the slideshow presentation, annotating slides, and creating sketches. Figure 3 shows our instructor's sketches illustrating how textures are mapped and how the perspective camera works.

When the instructor replays a recorded lecture, XRStudio will first load the video recordings and restore the scene content, then continuously load and replay the interaction recordings. This facilitates real-time playback of a class including any sketches that were created. Note that the instructor controlling the replay is still in the scene, while the instructor that previously recorded the lecture will appear as a ghost avatar. Our replay system makes it possible for the instructor to pre-record parts of their lecture, or load someone else's lecture and add to it, as a form of layered acting [10, 31].

Multiple Output Modalities *(R6).* Our descriptions of XRStudio have so far been from the instructor perspective. Students can watch a live stream or recorded lecture video, which are live-edited video productions based on the instructor's preset choices. The live stream from XRStudio can be accessed via videoconferencing software like Zoom or using the XRStudio student client. In the client, students can also join the lecture scene with an avatar, moving freely or switching to pre-calibrated camera perspectives. The student client supports 3D on desktop/tablet, common VR HMDs, and AR capable phone/tablet. Students can communicate with the instructor via XRStudio's built-in text, audio, and video chat.

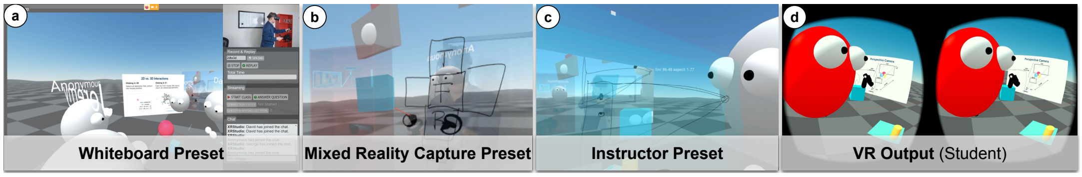

Figure 3: In our XRStudio system walkthrough, the instructor teaches a lesson on 2D vs. 3D design: *(a)* the class gathers around 3D objects—captured in third person from the perspective of the virtual whiteboard camera; *(b)* the instructor sketches out a texture and how it would be mapped to each side of the cube—captured using live compositing of the instructor's webcam and the calibrated virtual camera; *(c)* the instructor sketches the view frustum of a perspective camera from a student's point of view—captured from the instructor's first-person point of view; *(d)* the student observes the perspective rendering and clipping of a virtual cube as the instructor moves it closer—captured in first person from the student's stereoscopic view in VR.

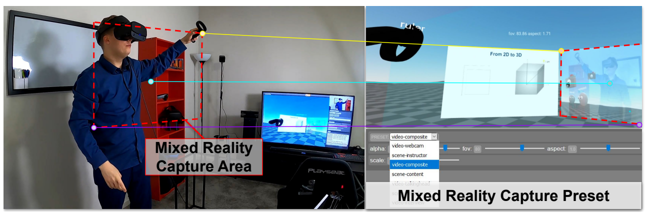

Figure 4: XRStudio implements a three-point calibration technique where the instructor first demonstrates to the system the location of the physical camera (cyan), then picks the top-left (yellow) and bottom-right (purple) corners to define the mixed reality capture area. This calibrates the position, rotation, FOV, and aspect ratio of the virtual camera. The instructor can tune each parameter so that VR representations of the controllers map the physical counterparts.

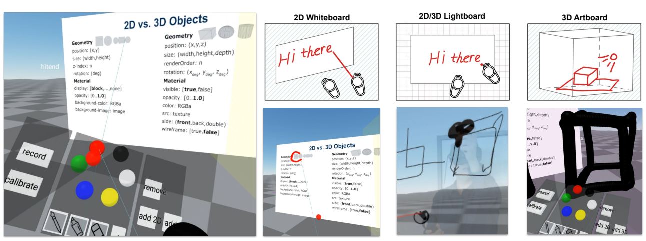

Figure 5: XRStudio supports three types of whiteboards for 2D/3D content creation in VR. Left: *2D whiteboards* for slides and 2D sketches. Middle: *2D/3D lightboards* for 2D/3D sketches in mixed reality capture. Right: *3D artboards* are containers for 3D sketches and 3D models.

Additional Cameras for Capture *(R1, R3, R6).* The system components we have described so far provide the core support for live streaming and record-and-replay of VR lectures. We implemented additional functionality in XRStudio to further demonstrate the potential of our system in more advanced scenarios.

To get instructors started, XRStudio provides a default scene with various camera angles following the instructor or showing close-ups of the whiteboard and 3D objects in the center of the scene (Figure 6a–b). For more advanced use cases, XRStudio implements a virtual camera management system, allowing instructors to create, position, and manipulate additional virtual cameras (Figure 6c). These cameras can be operated manually or set to automatically follow any 3D object in the scene including specific users. In the future, we could easily imagine pre-defined camera tracks to support captures as professional film studios might perform.

To connect additional cameras, we use a special XRStudio camera client that can be operated just like the student client in 3D, VR, or AR. The camera client supports mapping virtual cameras to physical cameras when paired with a webcam or an AR capable smartphone/tablet (Figure 6d). An experienced instructor could operate these cameras by themself, or they could work with a media designer to assist in the production of a live or recorded lecture. Finally, we also implemented a capture technique that streams the instructor's desktop into the 3D scene. This enables screen-cast with narration and allows the instructor to easily integrate any desktop or web application.

While we have tested these advanced features of XRStudio in our research team and also demonstrate them in the video accompanying the paper, we were not able to formally study them due to a combination of factors including COVID-19, which prevented the use of our university's VR equipped classrooms, and the limited access for students to VR and AR equipment.

This concludes our system walkthrough and rounds out the techniques we have available for instructors to make use of XRStudio's live production presets and mixed reality capture for VR lectures.

5.3 Implementation

XRStudio is a web application built on top of A-Frame such that only a modern browser with WebRTC support. Webcam access is required for mixed reality capture. The system implements the three-point calibration technique by projecting the webcam's live video feed to the instructor and prompting calibration points in VR, then maps the virtual camera frustum to the webcam. Person segmentation of the instructor is implemented using BodyPix[2]. Real-time synchronization of 3D content and interactions between the instructor and students is based on Networked A-Frame[3]. We implemented custom text, audio, and video chat components in A-Frame using WebRTC. Our record-and-replay system uses Firebase Realtime Database to keep all clients in sync.

More information about XRStudio and the source code are available at https://mi2lab.com/research/xrstudio.

[2]https://github.com/tensorflow/tfjs-models/tree/master/body-pix
[3]https://github.com/networked-aframe

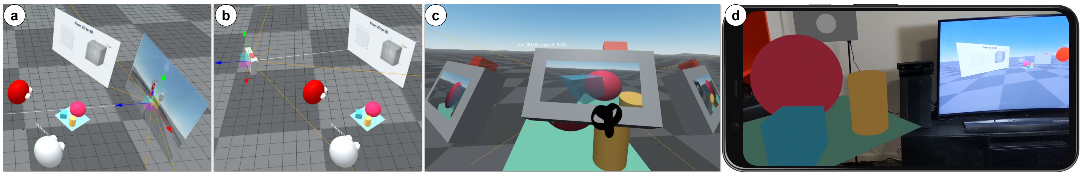

Figure 6: XRStudio supports multiple cameras in a scene which allows the capture from multiple angles, giving the instructor a lot of flexibility in how to present lecture content: *a)* the default camera for mixed reality capture; *b)* the default camera for whiteboard content; *c)* an instructor can place additional virtual cameras in VR; *d)* additional cameras can be added in AR, which may be operated by the instructor or a media designer.

6 WALKTHROUGH DEMONSTRATIONS

In the first part of our summative evaluation, we used XRStudio in workshops with the same five instructors interviewed earlier, conducting walkthrough demonstrations of XRStudio to critique the system from each instructor's perspective. As a common evaluation technique [28], we saw more value in conducting walkthrough demonstrations as opposed to usability studies, since we aimed to learn how instructors might use, and in fact wish to adapt, our tools for teaching courses in their own domains, while still gaining feedback on the utility of XRStudio's current features.

We conducted individual sessions with the instructors, presenting them with two lecture videos recorded with an implementation of XRStudio with two presets. The first video showed one member of our research team as the instructor, delivering a lecture in a virtual classroom using 3D sketching tools and virtual slides. The emphasis of this video was on the first-person instructor perspective. The second video showed a third-person recording of the instructor giving the same lecture using mixed reality capture. After each video, we engaged in a brainstorming and critique session, discussing how the instructors would change the classroom environment, content, and interactions in our example and probing into the affordances of mixed reality capture for their domains.

All instructors agreed that XRStudio could be useful for conducting live remote lectures with some customization of the the virtual environment and content, particularly when discussing the first-person lecture video. Some instructors expressed that they would walk students through the process of developing an immersive scene and have the environment update around them in real-time. There was also significant interest in conducting collaborative activities in XRStudio, with two instructors expressing that their "perfect tool" would allow for performing critiques of student work and reconfiguring the virtual setup directly in VR to mimic physical active learning environments. While all saw potential with this teaching style, some expressed that XR may not be necessary for concepts which do not involve 3D content and could be taught with just a slide deck. Additionally, there were concerns about the first-person perspective for lecturing, as some students had experienced nausea when watching recordings of VR experiences in past courses.

The use of mixed reality capture in the second lecture video sparked new ideas among the instructors beyond mimicking a physical classroom setup. Some expressed that mixed reality capture could be beneficial for live demonstrations that involve 3D sketching, visualizing complex concepts from different perspectives, and promoting spatial understanding, which is often difficult for students. The architecture instructor also felt it would be useful if the mixed reality capture features could extend to external drawing and modeling tools which support a higher level of detail and precision. Exploring the use of mixed reality capture to convey annotations of physical objects and locations rather than completely virtual environments seemed particularly promising to instructors. The nursing instructor proposed this for demonstrations of medical procedures, expressing that conducting an AR simulation where they are present in a real emergency room would enrich "the simulation in a way that is much more interactive." The instructor in landscape architecture expressed that during the current pandemic, mixed reality capture could be particularly powerful for "communicating existing places" and concepts for people who are unable to visit in person, rather than constructing new virtual sites.

7 END-TO-END LECTURE EXPERIENCE

In the second part of the summative evaluation, we used XRStudio to teach a lecture on 3D design, gathering feedback from three groups of students and the two media designers we had consulted.

7.1 Student Feedback

To gather feedback from students, we ran a remote study to compare three conditions: *slides*—a slideshow presented over Zoom with laser pointer; *instructor PoV*—the instructor's first-person VR view using XRStudio; and *MR capture*—live compositing using XRStudio's presets with an emphasis on mixed reality capture. Students could always see the raw webcam video of the instructor as well.

Procedure. We evaluated the three conditions in a between-subjects study in three sessions with 4–5 students each. The study began with a short survey about participant demographics and background experience with VR. Then, a live lecture on 3D design of approximately 8–10 minutes was given in each condition (Figure 7, left). We started each session with a different condition to minimize order effects. After each lecture, all students filled out a feedback survey where they were also asked to rank the lecture formats according to their preference. At the end of the session, we engaged in a brief discussion asking each student which lecture style they preferred overall and why they chose that style.

We recruited three groups of five Master's students from an introductory course on AR/VR technologies on campus (one no-show; $N = 14$, 7 male, 7 female). Participants were between 22 and 32 years old, and had a variety of majors: Information Science (6), HCI/ UX (5), Health Informatics (1), Landscape Architecture and Environmental Informatics (1), and Embedded Systems (1). Although most participants (12) had used a VR device at least once before, they overall had relatively little prior experience with VR technology (7 used VR 1-2 times in the past year, 2 used VR about every other month, 2 used VR about once a month, and 1 used VR about once a week).

Results. Overall, participants preferred both XRStudio conditions to the traditional lecture slides, with a slight preference for the *MR capture* condition over the *instructor PoV* (Figure 7). We used thematic coding of verbal and written survey responses to identify key findings.

Slides. Participants overwhelmingly rated *slides* as their least preferred condition. Participants attributed this, in part, to the static and non-interactive nature of delivery. Several participants additionally considered the 2D format of *slides* quite limiting for the 3D lecture content. One participant noted that trying to understand 3D rotations based on the slides "forces me to allocate thinking resources to trying to decipher how the 2D content translates into a 3D environment, while at the same time having to follow the lecture material."

Two of our 14 participants, however, preferred the *slides* condition. They agreed that while *slides* was not as engaging as the immersive conditions, it was nonetheless a "clear" method of conveying information. One participant said "it was easier for me to actually understand and let the information marinate because you were reading off of the slides," whereas in the immersive conditions, "I was so focused on the visuals it was a little distracting and some of the information didn't sink in." Several participants also found the static slide reference more conducive to note-taking.

Instructor PoV. Five of 14 participants preferred the *instructor PoV* condition. The immersive format using XRStudio was considered more engaging than the *slides* condition. In the words of one participant, "I felt like I was in the lecture rather than just watching it on Zoom." Participants also saw benefits of this immersive condition in supporting understanding of 3D content, whereas in the traditional slideshow, understanding 2D representations of 3D content required more thought. One additional feature of the *instructor PoV* condition which participants particularly appreciated was the visibility of 2D slides within the immersive environment. Participants found the slides as a helpful reference that complemented the instructor's interactive delivery of the lecture in 3D.

For many participants, the primary limitation of the *instructor PoV* condition was that it was uncomfortable and, to some, nauseating to watch. As noted by one participant, "it was a little disturbing that the screen was shaking constantly." One participant found the immersive interface "overwhelming," saying "it is hard for people who are new to [VR] to focus on what the prof is talking about." Several participants additionally disliked the instructor's use of free-hand drawing in 3D, finding the drawings disorganized and difficult to understand. Participants found the 3D drawings difficult to decipher from the background, some suggesting the use of a simpler virtual environment for visibility sake. Participants also suggested the use of pre-authored 3D content, stating that "unless the professor is a skilled artist and had very steady hands, 3D shapes such as cubes or rectangles would come out crooked."

MR Capture. Seven of 14 participants preferred the *MR capture* condition. *MR capture* shared many of the immersive instruction advantages with the *instructor PoV* condition, including greater student engagement and enabling interactions with the 3D content. The primary difference between the *instructor PoV* and *MR capture* conditions was the perspective. Participants found the third-person perspective of the MR view helpful as it allowed them to observe exactly how the instructor was interacting with content. One student said, "I liked being able to see [the professor] speak about 3D concepts while creating 3D objects. It allowed me to better understand what he was talking about." Participants also saw the third-person MR view as more stable and "natural" than the first-person VR view in the *instructor PoV* condition. Multiple participants compared the third-person perspective to watching the lecturer "in a real classroom" as opposed to to just a Zoom window. One said it was "more engaging since I felt like I was learning in a physical, hands-on environment. It fits more with my learning style."

While the instructor primarily used the third-person perspective in *MR capture*, in this condition, he had access to the full set of XRStudio presets, allowing him to switch the camera angle and capture modality. Participants appreciated that the instructor used the different viewpoints strategically as he was sketching, for example, how 3D transforms work and how textures are mapped to the faces of a cube. While many appreciated seeing this from both the first- and third-person perspective, switching presets was occasionally "distracting" and difficult to follow. As put by one participant, "I was a little confused when the screen changes from the presenter's perspective to another angle ... as the transition was fast." Participants who were new to VR generally had a harder time following the instructor, especially when switching perspectives, and attributed part of the confusion to a lack of experience with navigating virtual environments. A more consistent perspective could potentially help students focus on the content. According to one participant, with the *instructor PoV*, "for some students who don't have a VR headset, now they can have a really straightforward experience of using it, even through an instructor's eyes." Another minor concern voiced by several participants was that the presentation slides were not as visible as in the *instructor PoV* and *slides* conditions.

7.2 Media Designer Feedback

To elicit feedback from a media design perspective, we conducted a follow-up interview with the two professional media designers from our initial needfinding interviews. We sought their opinions comparing the same three lecture formats as above, *slides*, *instructor PoV*, and *MR capture*, discussing differences in affordances and production quality of lecture recordings from each condition.

Overall, the media designers preferred the *MR capture* recording, as it provided the most dynamic viewing experience without overwhelming the audience. They noted that the instructor was able to engage more directly with the lesson content, as opposed to the *slides* condition with a picture-in-picture view, where the instructor seemed more distanced from the content. Both media designers agreed that XRStudio's flexible camera configurations, particularly perspective switches, must be used carefully. Considering whether the utility of different viewpoints would generalize to other content, they emphasized the need to determine which lecture styles and perspectives are most effective on a case-by-case basis, by evaluating which of XRStudio's output modalities would best support

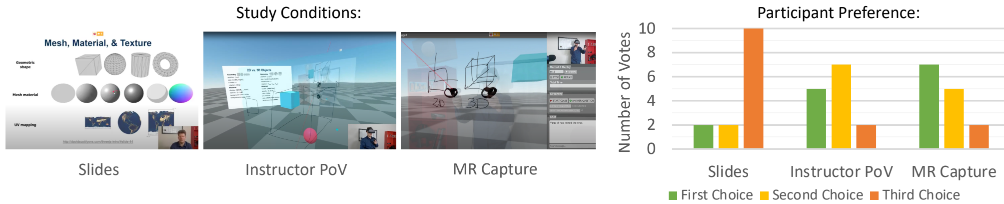

Figure 7: Left: The three conditions in our lecture evaluation study: *1) slides, 2) instructor PoV,* and *3) MR capture.* Right: Overall ranking of participant preferences. Note that the order of conditions was randomized and counter-balanced.

different types of lecture content. While they thought it feasible for instructors to use XRStudio independently without the need for a media team, they still recommended stylistic changes to improve the viewing experience for learners and provide instructors with consistent previews of the presets to help them make decisions around lecture styles and perspectives in VR.

8 DISCUSSION

In this section, we review the findings from our studies around XRStudio and discuss the limitations.

Good Lecture Topics. As hypothesized by us and instructors, we observed benefits of using XRStudio to teach 3D spatial concepts, so the *instructor PoV* and *MR capture* conditions were likely to be preferred. The "wow factor" with XR technology, as expressed in interviews, can further bias the results, so this promise may not hold for every topic. We suppose that complex concepts that can be expressed spatially and sequentially will benefit the most from XRStudio's mixed reality capture and record-and-replay system. As also argued by instructors, there is added value in using XRStudio for "communicating existing places," which can also be represented virtually and filled in using mixed reality capture when access is limited. XRStudio's presets give a lot of flexibility. An instructor can give a lecture mostly based on slides and switch to VR as necessary. The main benefit of using XRStudio is that its mixed reality capture modality makes it possible for students to perceive the VR experience via video, preserving the spatial relationship to the instructor, without having access to an HMD.

Instructor Effort. We characterize the effort required of instructors to make use of XRStudio as relatively minor, but there is still considerable technology involved. For our 2D vs. 3D lecture, we used the default scene and camera settings, and just imported slides. The main challenge was going through the three-point calibration, which simplifies the setup significantly, but still requires practice. Instructors experienced with XR technology may quickly reach the ceiling and look for additional features in XRStudio. Our interviewees appreciated that XRStudio can load common 3D file formats and any A-Frame scene as a starting point. This supports re-use and more complex lecture content. For example, the film instructor had pre-scripted animations in A-Frame, while the architecture and nursing instructors wanted to teach in custom environments, both of which are easily supported by XRStudio. Importing lecture content also reduces the need for sketching. It is possible to annotate slides or sketch around 3D objects in the virtual world or physical objects in the real world via mixed reality capture. Our architecture instructor who is used to working with high-poly models and high-precision modeling tools argued that our sketch annotation in XRStudio could be sufficient. He could always import higher quality assets from domain-specific tools if necessary.

Best Practices. While our three groups of stakeholders largely appreciated the dynamic use of presets (e.g., switching between first-person *instructor PoV* and third-person *MR capture*), some students argued they could not anticipate transitions, and therefore found the lecture difficult to follow, especially when preset transitions occurred too frequently. We therefore recommend that instructors consider the cognitive load associated with adjusting to new presets for viewers and prepare them for the transition. As the student feedback also suggests, preset transitions are most effective when used in a targeted manner. More experience teaching with XRStudio will allow an instructor to develop an intuition for when to use which preset and how to initiate transitions. It was promising to learn from media designers that just seeing our XRStudio recordings allowed them to recommend presets and transitions, which could further assist instructors in preparing the lecture material and making the best possible use of a system like XRStudio.

Additional Studies. Due to the COVID-19 pandemic, we were limited to conducting remote studies, and as such, could not test the XRStudio system with different instructors or with students directly joining the lecture in VR. To mitigate the first issue, we involved instructors in both the formative and summative evaluation stages through interviews and walkthrough demonstrations via Zoom. A study with students in VR is planned for the future when access to on-campus VR equipment is again possible. Our studies support the claims around mixed reality capture. The fact that XRStudio is available enables future studies exploring the usability and utility of systems like it in a variety of both virtual and real classroom environments with both instructors and students in VR.

9 CONCLUSION

We presented XRStudio, our system-driven exploration into how immersive content could be live streamed and benefit from virtual production techniques in filmmaking when adapted to the educational domain. We contribute insights on the potential of using mixed reality capture from evaluations with three key stakeholders: media designers, instructors, and students. We are excited about the potential these stakeholders commonly saw with XRStudio to positively impact teaching and learning given the new flexibility due to its mixed reality capture and record-and-replay system.

ACKNOWLEDGEMENTS

We thank U-M's XR Initiative and the Center for Academic Innovation for supporting this research.

REFERENCES

[1] Narges Ashtari, Andrea Bunt, Joanna McGrenere, Michael Nebeling, and Parmit K. Chilana. 2020. Creating Augmented and Virtual Reality Applications: Current Practices, Challenges, and Opportunities. In *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376722

[2] Jeremy N Bailenson, Nick Yee, Jim Blascovich, Andrew C Beall, Nicole Lundblad, and Michael Jin. 2008. The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences* 17, 1 (2008), 102–141.

[3] Catherine Ball and Kyle Johnsen. 2016. An accessible platform for everyday educational virtual reality. In *2nd IEEE Workshop on Everyday Virtual Reality, WEVR@VR 2016, Greenville, SC, USA, March 20, 2016*. IEEE Computer Society, New York, NY, USA, 26–31. https://doi.org/10.1109/WEVR.2016.7859540

[4] Liwei Chan and Kouta Minamizawa. 2017. FrontFace: facilitating communication between HMD users and outsiders using front-facing-screen HMDs. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2017, Vienna, Austria, September 4-7, 2017*, Matt Jones, Manfred Tscheligi, Yvonne Rogers, and Roderick Murray-Smith (Eds.). ACM, New York, NY, USA, 22:1–22:5. https://doi.org/10.1145/3098279.3098548

[5] Chih-Ming Chen and Chung-Hsin Wu. 2015. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education* 80 (2015), 108–121.

[6] Jiawen Chen, Shahram Izadi, and Andrew W. Fitzgibbon. 2012. KinÊtre: animating the world with the human body. In *The 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, Cambridge, MA, USA, October 7-10, 2012*, Rob Miller, Hrvoje Benko, and Celine Latulipe (Eds.). ACM, New York, NY, USA, 435–444. https://doi.org/10.1145/2380116.2380171

[7] Alan Cheng, Lei Yang, and Erik Andersen. 2017. Teaching Language and Culture with a Virtual Reality Game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017*, Gloria Mark, Susan R. Fussell, Cliff Lampe, m. c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, New York, NY, USA, 541–549. https://doi.org/10.1145/3025453.3025857

[8] Kevin Chow, Caitlin Coyiuto, Cuong Nguyen, and Dongwook Yoon. 2019. Challenges and Design Considerations for Multimodal Asynchronous Collaboration in VR. *Proc. ACM Hum. Comput. Interact.* 3, CSCW (2019), 40:1–40:24. https://doi.org/10.1145/3359142

[9] Charles Crook and Louise Schofield. 2017. The video lecture. *The Internet and Higher Education* 34 (2017), 56–64.

[10] Mira Dontcheva, Gary D. Yngve, and Zoran Popovic. 2003. Layered acting for character animation. *ACM Trans. Graph.* 22, 3 (2003), 409–416. https://doi.org/10.1145/882262.882285

[11] Thierry Duval and Cédric Fleury. 2009. An asymmetric 2D Pointer/3D Ray for 3D interaction within collaborative virtual environments. In *Proceeding of the 14th International Conference on 3D Web Technology, Web3D 2009, Darmstadt, Germany, June 16-17, 2009*, Stephen N. Spencer, Dieter W. Fellner, Johannes Behr, and Krzysztof Walczak (Eds.). ACM, New York, NY, USA, 33–41. https://doi.org/10.1145/1559764.1559769

[12] Daniel Fällman, Anders Backman, and Kenneth Holmlund. 1999. VR in education: An introduction to multisensory constructivist learning environments. In *Conference on University Pedagogy, Umea University, Umea, Sweden, February 18-19, 1999*. Umea University, Umea, Sweden, 10.

[13] Dustin E. R. Freeman, Stephanie Santosa, Fanny Chevalier, Ravin Balakrishnan, and Karan Singh. 2014. LACES: live authoring through compositing and editing of streaming video. In *CHI Conference on Human Factors in Computing Systems, CHI'14, Toronto, ON, Canada - April 26 - May 01, 2014*, Matt Jones, Philippe A. Palanque, Albrecht Schmidt, and Tovi Grossman (Eds.). ACM, New York, NY, USA, 1207–1216. https://doi.org/10.1145/2556288.2557304

[14] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017*, Gloria Mark, Susan R. Fussell, Cliff Lampe, m. c. schraefel, Juan Pablo Hourcade, Caroline Appert, and Daniel Wigdor (Eds.). ACM, New York, NY, USA, 4021–4033. https://doi.org/10.1145/3025453.3025683

[15] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, New York, NY, USA, 54. https://doi.org/10.1145/3173574.3173628

[16] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How video production affects student engagement: an empirical study of MOOC videos. In *First (2014) ACM Conference on Learning @ Scale, L@S 2014, Atlanta, GA, USA, March 4-5, 2014*, Mehran Sahami, Armando Fox, Marti A. Hearst, and Michelene T. H. Chi (Eds.). ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2556325.2566239

[17] Anna Hansch, Lisa Hillers, Katherine McConachie, Christopher Newman, Thomas Schildhauer, and J. Philipp Schmidt. 2015. Video and online learning: Critical reflections and findings from the field. *HIIG Discussion Paper Series* 2 (2015), 1–34.

[18] John Hedberg and Shirley Alexander. 1994. Virtual Reality in Education: Defining Researchable Issues. *Educational Media International* 31, 4 (1994), 214–220. https://doi.org/10.1080/0952398940310402 arXiv:https://doi.org/10.1080/0952398940310402

[19] Robert Held, Ankit Gupta, Brian Curless, and Maneesh Agrawala. 2012. 3D puppetry: a kinect-based interface for 3D animation. In *The 25th Annual ACM Symposium on User Interface Software and Technology, UIST '12, Cambridge, MA, USA, October 7-10, 2012*, Rob Miller, Hrvoje Benko, and Celine Latulipe (Eds.). ACM, New York, NY, USA, 423–434. https://doi.org/10.1145/2380116.2380170

[20] Sandra Helsel. 1992. Virtual Reality and Education. *Educational Technology* 32, 5 (1992), 38–42. http://www.jstor.org/stable/44425644

[21] Roland Holm, Erwin Stauder, Roland R. Wagner, Markus Priglinger, and Jens Volkert. 2002. A Combined Immersive and Desktop Authoring Tool for Virtual Environments. In *IEEE Virtual Reality Conference 2002, VR '02, Orlando, Florida, USA, March 24-28, 2002, Proceedings*. IEEE Computer Society, New York, NY, USA, 93–100. https://doi.org/10.1109/VR.2002.996511

[22] Akira Ishii, Masaya Tsuruta, Ippei Suzuki, Shuta Nakamae, Tatsuya Minagawa, Junichi Suzuki, and Yoichi Ochiai. 2017. ReverseCAVE: providing reverse perspectives for sharing VR experience. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference, SIGGRAPH '17, Los Angeles, CA, USA, July 30 - August 03, 2017, Posters Proceedings*. ACM, New York, NY, USA, 28:1–28:2. https://doi.org/10.1145/3102163.3102208

[23] G. Michael Lemole Jr., P. Pat Banerjee, Cristian Luciano, Sergey Neckrysh, and Fady T. Charbel. 2007. Virtual Reality in Neurosurgical Education: Part-task Ventriculostomy Simulation with Dynamic Visual and Haptic Feedback. *Neurosurgery* 61, 1 (07 2007), 142–149. https://doi.org/10.1227/01.neu.0000279734.22931.21 arXiv:https://academic.oup.com/neurosurgery/article-pdf/61/1/142/24538719/00006123-200707000-00016.pdf

[24] Hannes Kaufmann, Dieter Schmalstieg, and Michael G. Wagner. 2000. Construct3D: A Virtual Reality Application for Mathematics and Geometry Education. *Educ. Inf. Technol.* 5, 4 (2000), 263–276. https://doi.org/10.1023/A%3A1012049406877

[25] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George W. Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, UIST 2019, New Orleans, LA, USA, October 20-23, 2019*, François Guimbretière, Michael Bernstein, and Katharina Reinecke (Eds.). ACM, New York, NY, USA, 161–174. https://doi.org/10.1145/3332165.3347872

[26] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. 2019. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 284. https://doi.org/10.1145/3290605.3300514

[27] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. 2020. TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators. In *UIST '20: The 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, October 20-23, 2020*, Shamsi T. Iqbal, Karon E. MacLean, Fanny Chevalier, and Stefanie Mueller (Eds.). ACM, New York, NY, USA, 182–195. https://doi.org/10.1145/3379337.3415827

[28] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation Strategies for HCI Toolkit Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*, Regan L. Mandryk, Mark Hancock, Mark Perry, and Anna L. Cox (Eds.). ACM, New York, NY, USA, 36. https://doi.org/10.1145/3173574.3173610

[29] Stefan Marks, David White, and Manpreet Singh. 2017. Getting up Your Nose: A Virtual Reality Education Tool for Nasal Cavity Anatomy. In *SIGGRAPH Asia 2017 Symposium on Education* (Bangkok, Thailand) *(SA '17)*. ACM, New York, NY, USA, Article 1, 7 pages. https://doi.org/10.1145/3134368.3139218

[30] Zahira Merchant, Ernest T. Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J. Davis. 2014. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Comput. Educ.* 70 (2014), 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

[31] Michael Nebeling, Katy Lewis, Yu-Cheng Chang, Lihan Zhu, Michelle Chung, Piaoyang Wang, and Janet Nebeling. 2020. XRDirector: A Role-Based Collaborative Immersive Authoring System. In *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy

Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik (Eds.). ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376637

[32] Michael Nebeling and Katy Madier. 2019. 360proto: Making Interactive Virtual Reality & Augmented Reality Prototypes from Paper. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 596. https://doi.org/10.1145/3290605.3300826

[33] Sebastian Oberdörfer, David Heidrich, and Marc Erich Latoschik. 2019. Usability of Gamified Knowledge Learning in VR and Desktop-3D. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 175. https://doi.org/10.1145/3290605.3300405

[34] Joseph Psotka. 1995. Immersive training systems: Virtual reality and education and training. *Instructional Science* 23, 5/6 (1995), 405–431. http://www.jstor.org/stable/23370939

[35] Muhamad Izzat Rahim and Sarimah Shamsudin. 2019. Video Lecture Styles in MOOCs by Malaysian Polytechnics. In *Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology* (Nagoya, Japan) *(ICEMT 2019)*. ACM, New York, NY, USA, 64–68. https://doi.org/10.1145/3345120.3345169

[36] George G. Robertson, Stuart K. Card, and Jock D. Mackinlay. 1993. Three views of virtual reality: nonimmersive virtual reality. *Computer* 26, 2 (1993), 81–. https://doi.org/10.1109/2.192002

[37] José Miguel Santos-Espino, María Dolores Afonso-Suárez, and Cayetano Guerra-Artal. 2016. Speakers and boards: A survey of instructional video styles in MOOCs. *Technical Communication* 63, 2 (2016), 101–115.

[38] Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, and Wilmot Li. 2019. Interactive Body-Driven Graphics for Augmented Video Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos (Eds.). ACM, New York, NY, USA, 622. https://doi.org/10.1145/3290605.3300852

[39] Christopher D. Wickens. 1992. Virtual reality and education. In *Proceedings of the 1992 IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 1. IEEE, New York, NY, USA, 842–847. https://doi.org/10.1109/ICSMC.1992.271688

[40] William Winn. 1993. A conceptual basis for educational applications of virtual reality.

[41] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. 2018. Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality. In *The 31st Annual ACM Symposium on User Interface Software and Technology, UIST 2018, Berlin, Germany, October 14-17, 2018*, Patrick Baudisch, Albrecht Schmidt, and Andy Wilson (Eds.). ACM, New York, NY, USA, 853–866. https://doi.org/10.1145/3242587.3242597

[42] Kexin Yang, Xiaofei Zhou, and Iulian Radu. 2020. XR-Ed Framework: Designing Instruction-driven and Learner-centered Extended Reality Systems for Education. *CoRR* abs/2010.13779 (2020), 1–21. arXiv:2010.13779 https://arxiv.org/abs/2010.13779

[43] Yu Zhu, Kang Zhu, Qiang Fu, Xilin Chen, Huixing Gong, and Jingyi Yu. 2016. SAVE: shared augmented virtual environment for real-time mixed reality applications. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry, VRCAI 2016, Zhuhai, China, December 3-4, 2016*, Yiyu Cai and Daniel Thalmann (Eds.). ACM, New York, NY, USA, 13–21. https://doi.org/10.1145/3013971.3013979