3D VR180 Livestreaming System for Semi-Synchronous Distance Learning in TEFL Applications

Steven H. Urueta (surueta@keio.jp)
Graduate School of System Design and Management, Keio University, Japan
Department of English Language Studies, Mejiro University, Japan

Tetsuro Ogi (ogi@sdm.keio.ac.jp)
Graduate School of System Design and Management, Keio University, Japan

Abstract

VR video formats designed for use with head-mounted displays (HMDs) can be useful for immersive learning and allow students to freely focus their gaze on objects or people of interest. However, very little research has been done on the use of such technology as a medium- or long-term substitute for traditional classroom settings. This is particularly evident in the TEFL (Teaching English as a Foreign Language) field. To address the issue of providing an immersive, collaborative active learning environment while maintaining the privacy of individual students, a livestreaming system was created using 3D VR180 video (a stereoscopic video format that allows an HMD-wearing user to experience a 180-degree field of view). This system allows students to anonymously communicate with both the lecturer and one another on a message board using speech recognition software. To assess the potential of this system for expanded use, a mixed-methods approach involving qualitative user studies and quantitative benchmarking was undertaken.

Keywords: Virtual Reality, Computer-Assisted Language Learning, Teaching English as a Foreign Language, Head-Mounted Display

Introduction

Virtual reality video formats viewed through head-mounted displays (HMDs) may offer a high level of immersion. More specifically, VR180 video allows for turning one's head to gaze at objects of interest within a wide field of view, 180 degrees, as well as experience a video stream in stereoscopic 3D (Cooper et al., 2019). This can mean an increased feeling of being present at the filming location, an example of a phenomenon known as "telepresence", which describes the use of various technologies to create the effect of being at a different location. This has also explored in past research related to the use of VR headsets in education specifically (Guevara et al., 2020). However, the applications of such technology in teaching English as a foreign language (TEFL), especially when used in a capacity to receive a live video feed in real time, are understudied (Radianti et al., 2020). This project explores the possibilities and limitations of distance learning in a TEFL context based on live VR180 video. This

includes as a substitute for a more traditional classroom experience or distance learning experience, as well as how the results open the door for interdisciplinary research. It draws from the fields of Computer-Assisted Language Learning, Human-Interface Interaction, and Educational Technology.

To better understand these intersections, as well as to address a demand for immersive learning by students, a VR livestreaming system was built and accompanying lessons for it were created. The livestreaming system allows for an instructor to broadcast their classroom in stereoscopic VR to multiple HMD-wearing students simultaneously. In this system, students can communicate with the teacher and one another using a message board controlled by voice commands. Through these commands, students trigger the recording and submission of their spoken comments, which are transcribed into text. This transcription is handled by AI-enhanced speech-to-text technology. As students engage with the lesson, the teacher can receive feedback through the message board to adjust the pacing or content, as well as respond directly to questions. As the time lag for such communication is longer than ordinarily found for synchronous learning but much smaller than that for asynchronous learning, this system has been dubbed as "semi-synchronous."

The semi-synchronous state brought about by using this message board results in two characteristics. The first is that the time lag found with many VR180 streaming services, which can be upwards of 30 seconds, is no longer a fundamental obstacle to two-way communication. The second is that it may provide an immersive, collaborative learning environment for students that wish to maintain privacy online by not displaying themselves visually or auditorily.

Five lessons were planned specifically for VR instruction through this system, following tenets of active learning and task-based learning (Lee, 2016; Lutes, 2018), and such use was also accounted for in the planning stage of the system design process itself. These lessons were used as a substitute for a portion of a 10-week university skill-building English course targeted towards science and engineering students. In this course, which was originally taught in person and adapted more recently for a video chat platform (Zoom), one of the major goals was for students to be able to describe and give queries about the appearance and function of items. This was considered a necessary skill for participating in conferences and explaining and understanding how various equipment works.

This system will be examined to help answer three main questions. The first is, "Is it possible to integrate active-learning tenets in a medium-term course with a system exclusively using VR?" The second is, "Is it possible to create a reliable system that can be used in common student use cases?" The third is "What do recent technological developments, including the results of this study, mean for the adoption of VR in CALL, as well as the relationship between CALL and VR system design?" This includes answering questions about forms of privacy in VR worlds and scenarios, lesson planning for teachers conducting classes in this new format, issues and problems that may appear in similar systems, and how these technologies may be taken advantage of to a greater degree.

To answer the first question, a small-n, five-week study was undertaken with two students. To assess the potential of this system for expanded use, a mixed-methods approach involving surveys, tests, and a qualitative user study was undertaken. To answer the second question, on system reliability, a series of tests related to latency,

speech-to-text accuracy, and voice macro accuracy were conducted. To answer the third question, other possible features for long-term courses were also tested and explored.

Previous Literature

According to Freina and Ott (2015), there are two main types of virtual reality. The first is "non-immersive" virtual reality, which uses ordinary computer equipment (such as a desktop PC or tablet with a screen) to display content. "Immersive" virtual reality, on the other hand, requires specialized equipment designed for a heightened sense of immersion, or the perception of being present in a non-physical world. This is often accomplished through an HMD or CAVE system (Cave Automatic Virtual Environment).

CAVE systems are a walk-in environment surrounded by screens, usually projected from the rear, with a user wearing specialized shutter glasses that allow for 3D viewing. A motion capture system allows for adjusting the viewpoint based on the user's position (Defanti et al., 2010). An example of CAVE use is the immersive examination of a 3D object from multiple angles.

HMDs, known to many consumers as "VR Goggles" or "VR Headsets", consist of images displayed on a screen or pair of screens that are mounted within a helmet or glasses. Tracking sensors tell a computer where the user is looking, and the computer then shows that point of view to the user. Practically, this means that a user can look at a computer-generated world (or pre-recorded 3D video footage) in a manner similar to being in the real world (Sherman & Craig, 2020).

A study by Jensen et al. (2018) found that VR technologies were not universally useful for skills acquisition. In some cases, they were counterproductive when compared with traditional teaching methods or digital environments with less immersion, such as a 2D display and personal computer. Rather, VR was most effective for a narrower band of situations requiring the processing of spatial or visual knowledge. "psychomotor skills related to head movement, such as visual scanning or observational skills." This was taken into account when designing the course.

This research focuses on the use of VR video, of which there two main types, VR180 and VR360. VR360 video is an older format, which allows a viewer to pan 360 degrees to view any desired portion of the video feed. VR180 is a newer format popularized by Lenovo and Google, which decreases the field of view to 180 degrees but has lower hardware and bandwidth requirements. Alternatively, it can have higher resolution for a given bandwidth. For distance learning with HMDs, current TEFL research mainly examines 360-degree video without 3D stereoscopic feeds or involves the short-term use of simple "Google-cardboard"-style devices as a supplement to a traditional classroom (Freina & Ott, 2015). Most of the content is pre-recorded, and papers dealing with active learning with low-latency, high-bitrate HMDs are mostly theoretical, especially so for live streams. Finally, the use of VR is often a one-time experience added as a supplement to an existing face-to-face course.

On the other hand, Yildirim, G. et al. (2019) discovered that while VR videos may not have greater short-term learning outcomes compared to 2D video, long-term outcomes can be greater. Given that the skills of object and function description are useful in a long-term sense for many students, this may hold promise.

On the content production front, it has been pointed out that many existing 360-degree video production workflows are in need of streamlining, often requiring post processing, manual uploading, and other steps (Feurstein, 2018). This may lead to a high barrier for teachers aiming to implement it into their courses, let alone design a course to be conducted for long periods and/or exclusively with VR video.

There are various recommendations towards the recording of educational video in a VR setting, including preventing distractions from the surroundings, preventing motion of the camera, and positioning the camera for capturing detail (Kavanagh, S. et al., 2016). Welking et al. further classify recording approaches to 360-degree videos, dividing them into the *didactical approach, preparation of the presenter, technical requirements, location and positioning, and recording process* (2019). However, there is little information specific to TEFL and how these recommendations apply to VR180 video requires further exploration.

While livestreaming video is as old as live television broadcasting, and larger-scale educational streams over the internet have been growing rapidly (Chen et al., 2021), the livestreaming of VR video is a far more recent development. While the idea of semi-synchronous learning using online tools is well-established, even in the TEFL field (Matsuura et al., 2004; Lanvin & Beaufait, 2003), a clear idea of the term "semi-synchronous" has not been formed for VR use cases, though semi-synchronous lecture platforms have been shown to allow for high levels of student collaboration and interaction (Kutnick & Joyner, 2019).

Finally, on the issue of privacy, maintaining it during e-learning situations has been shown to be important for students (Chen et al., 2021), including those in China and Japan (Yang & Wang 2014). Particularly with distance learning video chat applications such as Zoom, there have been concerns over the display of student information and likenesses, as well as leaks of data, including personal information (Fudge & Williams, 2020).

System Architecture and Explanation

Following an interdisciplinary approach, this research melds VR technology with TEFL tenets to create a livestreaming system in which immersive, live instruction can be provided to students. The system requirements were made and implemented following a system design perspective, with consultations undertaken with both students and teachers. Some of these requirements include allowing students to participate in the lesson without having to remove their headset, being able to be operate the system hands-free, the ability to view items in three dimensions, and allowing for two-way communication between the teacher and students while maintaining student privacy. Apart from immersiveness, the issue of privacy was paramount to a few students, as university policies sometimes required the recording and storage of all distance learning lessons. These students did not wish to have their likenesses accessible online, regardless of data-protection methods.

Figure 1
A diagram showing the architecture of the livestreaming system created for this research study. The arrows and lines represent commands and data flows within the system
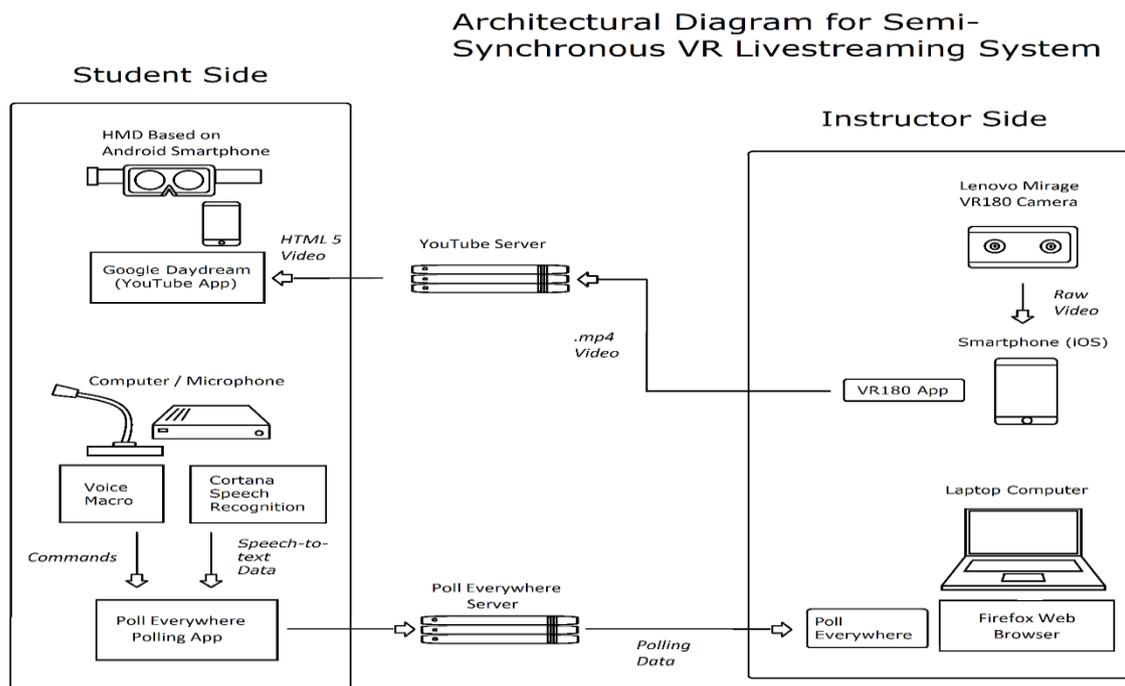
Student Side

Figure 1 shows an overview of the system. The left-hand box represents the equipment in the student's location. The student wears a Google Daydream HMD, which holds an Android Smartphone (Samsung Galaxy S9). The display of the smartphone, when the phone is inserted into the HMD, acts as the HMD's display. This is a common setup found in the popular "Google Cardboard" headsets and most cell-phone compatible HMDs. However, the smartphone is one of the few compatible with "Google Daydream", an app and headset combination that provides the ability to view YouTube videos in VR. Compared to many other cell phone-based systems, the Daydream's compatibility requirements mean a much higher level of image resolution, wider field of view, and a lower latency when moving the headset. These all may lead to an increase in telepresence compared to the more basic "cardboard" headsets or universal headsets. With this headset, worn with a strap for hands-free operation, the student views 3D VR180 content live as streamed by the instructor. The above configuration was chosen to minimize costs for student adoption while maintaining a baseline of performance, as well as keeping low barriers for setup and operation by instructors. While the latest desktop-based HMD systems, such as the Oculus Rift S and Valve Index, provide higher resolutions, frame rates, and responsiveness to motion inputs, a powerful PC and a relatively complicated setup procedure is required.

For the student to communicate through the message board, a personal computer running Windows (NEC Lavie Note Mobile / HP Omen 15) was loaded with three pieces of software. The first is "Voice Macro", a macro controller that uses voice commands to control programs running on the OS. The second is Microsoft's Cortana speech recognition engine, which uses adaptive learning and cloud-based processing to convert students' dictation into text. The third is the web-based app of "Poll Everywhere", a service that allows users to enter interactive meetings. Students use the Poll Everywhere application's message board, where the text versions of their dictated comments are submitted. These submissions are triggered by voice macro recognition, so a keyboard, mouse, or other handheld input device are unnecessary.

An example communication flow is as follows. Two main commands for Voice Macro, when spoken by the student, control sending comments to the message board. The "Answer Question" command opens the Polling Everywhere messaging window, activates the speech-to-text software, and begins the speech recognition process. The student then speaks his or her question or comment aloud. When finished, the student gives the command, "Stop Dictating." This command stops Cortana's recognition and instructs Polling Everywhere to send the message data to the Poll Everywhere server.

One of two connection points between the student and instructor sides is the Poll Everywhere server. This server receives student messages and sends them to the instructor side through the internet. As can be seen in the diagram, the portions of the system used for VR video streaming and for messaging are separate.

Instructor Side

On the instructor side, the instructor is in a physical classroom space, not wearing a VR headset. A VR180 camera (Lenovo Mirage) mounted on a tripod records the instructor's lesson. Spatial audio of the lesson is captured through the camera's two built-in microphones, and two fish-eye lenses capture the wide field of view necessary for VR180 recording. The contents of the lesson, position of the camera and teacher, etc. are outlined in Section 3.1, as they varied by lesson. The VR180 camera continuously sends the video data via a wireless connection (Wi-Fi Direct) to an iOS smartphone. The phone, running Google's VR180 app, uploads the live video stream to YouTube's servers, which can then be viewed live by the student(s).

During the lesson, the instructor mainly interfaces with a laptop computer. This computer displays the web application of Poll Everywhere; more specifically, its online message board. The teacher can view student comments as they enter the message board. In addition, one portion of the research study sessions was conducted with a projector displaying student comments for all participants to see.

Lesson Contents

Five lessons were given, revolving around the goal of describing the appearance and functions of unknown items. Following task-based learning and active learning tenets, as outlined by Ellis (2003, 2009), students explored ways to describe motion and manipulation. The unknown items, referred to as "mysterious items", were generally unidentifiable by the students. The aim of this was to ease students into unfamiliar territory, as they would have no direct basis of comparison in their first language.

The impetus behind the lesson design came from past assignments for the university course. For example, in a different class, students had the option of designing their own inventions. In face-to-face sessions, those that had done so had the opportunity to explain the inventions they designed to others, as well as attempt to understand the purpose and functions of the inventions their classmates designed. In these lessons, students would learn useful expressions and vocabulary, both technical and those found in everyday use, relating to describing the functions of these items. As the students involved had science- and engineering-related backgrounds, and the course focused on the use of English in a technical capacity, these skills fit hand-in-hand with the course's content. However, in a distance learning capacity, when the course was taught through Zoom, a few students complained about a disconnect from immersion. As a response to that, this VR activity was designed to build similar skills.

**Table 1**
Sample Mysterious Items

Item Name and Picture	Related Phrases and Concepts
Corn Kernel Remover (FVFTK-01 Corn Cutter by Shimomura Cutlery Co., Ltd.)	"Wrapping around" "Pulling through" "Adjustable" "Ergonomic" "Gripping" "Adjusting"
Stress Relief Ball ("Mesh Squeeze Ball" by Top-Ace Co., Ltd.)	"Wrapping" "Bulging" "Absorbing" "Dimpling" "Netting" "Securing"

Table 1 above shows two of the actual mysterious items and related phrases and concepts that were taught. A total of eight mysterious items were presented over the five lessons. For two lessons, one mysterious item was shown, and for three lessons, two were shown.

The lesson would generally begin with the mysterious item being introduced visually. The item would be shown at various angles by the instructor, placed on a rotating stand, or displayed with a combination of both methods. Students were first tasked with guessing the functions of the items using their own vocabulary.

Figure 2

A display of how an HMD-wearing student may shift his or her area of focus while watching the lesson. Video snapshots taken with the Lenovo Mirage VR180 Camera

Raw Video Feed   Student Focus on Message Board   Student Focus on Instructor / Item
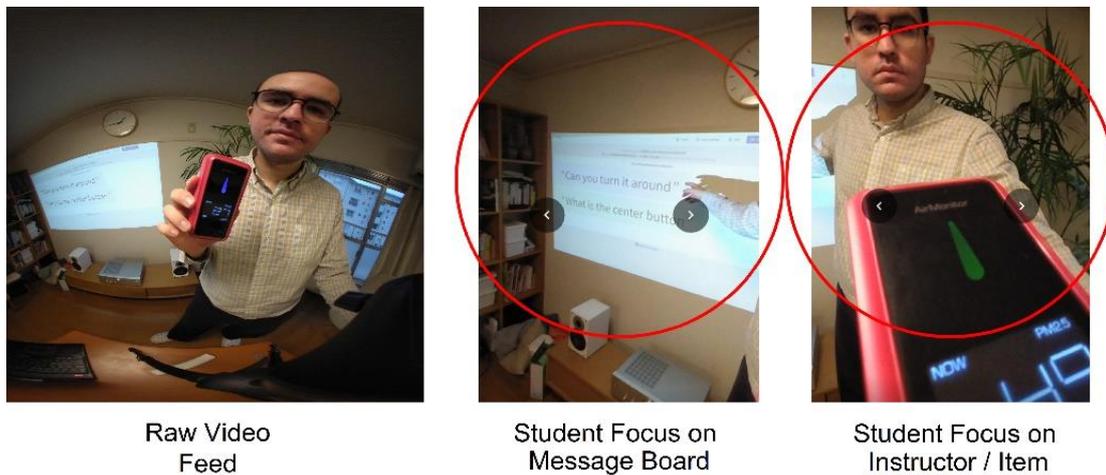
In Figure 2 above, the student's view of the lesson area shows how the student can focus on different parts of the lesson. On the left-hand side is the raw video feed. Only one half of the stereoscopic view is shown in the figure, and fisheye distortion is apparent. The student can see only a portion of the feed at a time and can control that portion by moving his or her head. With the two right-hand pictures, which are corrected for distortion and field of view, it is apparent that the student's gaze can shift from the message board to the object to the instructor freely. The red circle approximates the student's viewable area of focus. In Figure 2, the "mysterious item" in question is a PM 2.5 air quality sampler.

In the next part of the lesson, the instructor would provide a series of hints (such as where the item might be used, or who might use it), accompanied by physical examples of the items in use. The students would continue to make inferences using their own vocabulary, involving the "reasoning-gap" task type, which necessitates deriving new information (the purpose of the item) from given information (visual representation of the object, and clues as to its use) (Kozlova, 2018). Finally, methods to describe complicated forms of motion and manipulation were taught as the students experimented with their descriptions. For three of the lessons, this was repeated with a second mysterious item.

Table 2

Feedback and Resulting Iterative Changes

Lesson Number	Feedback	Iterative Changes
1	Speech and macro recognition errors were vast and constant.	Intensive calibration of voice/macro recognition was conducted as well as

		rechecked before subsequent lessons.
2	Difficulty in focusing on the teacher and on details of the items.	The distance between the camera and teacher/items was reduced.
3	Students had no way to communicate with one another* or verify the accuracy of their messages.	The virtual chatroom was projected behind the instructor, viewable by all participants.
4	Students had asked for written explanations of some phrases.	The projected virtual chatroom window became switchable with one of a pen-tablet-powered virtual whiteboard.
5	Students hoped for new types of learning methods, as well as being able to send audio recordings.	N/A (Final Lesson)

Source: Feedback from Student A and Student B during semi-structured interviews. One portion of external feedback after Lesson 3 was used for an iterative change (displayed with an asterisk).

Based on the feedback from the students during semi-structured interviews after each lesson (as well as a suggestion from another instructor), changes were implemented in later lessons, as can be seen in Table 2. Two pieces of equipment, an LED projector and pressure-sensitive writing tablet, were later added. While the final lesson still resulted in feedback for improvement, the lack of a subsequent lesson meant it could not be attempted, though it will be considered for future research.

As an aside, the first two lessons were conducted separately for each student. From the third lesson on, both students participated simultaneously.

Research Methodology and Study Details

Methodology

The study is divided two main parts. The first was a mixed methods student outcome evaluation using semi-structured interviews, surveys, and pre- and post-tests. It sought to understand difficulties faced by the students and how to deal with them, whether students had trouble learning, and possibilities for future improvement. It sought to answer the research question earlier mentioned: "Is it is "possible to integrate active-learning tenets in a medium-term course with a system exclusively using VR?"

The second was an evaluation of system performance, including quality of speech recognition, reliability of macro use, video lag, and internet bandwidth throughput. During the first portion, students were briefed on the system and its controls,

and they later underwent a series of exercises to improve the system's responsiveness, particularly with speech recognition and macro commands. The question to be answered here is, "Is it possible to create a reliable system that could be used in common student use cases?"

The results exploring the first two questions, as well as an overview of recent technological developments in VR and livestreaming, seek to answer the third question: "What do recent technological developments, including the results of this study, mean for the adoption of VR in CALL, as well as the relationship between CALL and VR system design?"

Student Experience and Outcome Evaluation

The mixed-methods evaluation began with pre- and post-testing, though starting only from the third lesson and ending in the fifth. Following Karshmer and Bryan's conception of a "one-shot" pre-test and post-test (Bryan & Karshmer, 2013), a ten-question test was given before and after the lesson. Each question focused on a learning concept as explored in the lesson. Students could select from four choices, as well as being able to respond that they did not know the answer. This was made in the hopes of reducing the effect of guessing, following a method outlined by Burton (2001).

Figure 3

A sample page from a pre- and post-test created for this study

A usability and learning survey was also given, this one for the final four lessons. Here, multiple-choice questions about student experiences were answered. A portion of them were useful for the student outcome evaluation, namely those related to excitement, engagement, and continued interest in learning. Another portion of the survey, related to lag, audio quality, video quality, and so on, reflected the students' assessments of their experience. Finally, a free-answer portion of each survey allowed students to respond about any likes or dislikes related to the study.

The semi-structured interviews followed a mix of closed- and open-ended questions, including follow-up "Why?" and "How?" questions that expanded on the students' observations. Adams (2015) states that semi-structured interviews can be especially useful for examining "uncharted territory with unknown but potential momentous issues" that require interviewers to have "maximum latitude to spot useful leads and pursue them." As this is a modular system that requires active adjustment over the course of instruction, and many large issues, particularly with module compatibility, may arise, such an approach was chosen. The semi-structured interviews were between 10-25 minutes long and conducted immediately after each of the final four sessions, following the ideas of probing and follow-up (Drever, 2003). Sample probing and follow-up questions can be found in Table 2 below, and, following SSI practices, a portion of the follow-up questions were not decided beforehand, and the interview contents were coded to find overarching themes.

Table 3
Sample Probing and Follow-up Questions

What was one of your most valuable learning experiences during this lesson?	In what way was this valuable?
How do you feel about the interaction you shared with the other student?	Would this apply if multiple students joined the class?
How do you feel about the interaction you shared with the teacher?	Why would more direct interaction be desirable?
How helpful were instructions given before the lesson (on how to conduct the lesson)?	How could the instructions be improved?
How has your experience been with the text-to-speech function?	How long would you be willing to optimize the text-to-speech quality through practice sessions?
Did you feel this lesson impacted your interest in learning English?	In what way was it impacted?
What challenges, positive and negative, did you face while taking this lesson?	Could you expand on how to tackle this negative challenge?

Finally, the two students already had experience with VR and had shown no signs of VR sickness or other discomfort within that experience.

System Performance Evaluation

The quality of speech recognition was measured through a commonly accepted metric, the Word Error Rate, in percent (WER%). In addition, the rate of correct macro

activations was also measured. To measure the WER, transcripts of the recordings of the students' speech were compared with the messages uploaded to Poll Everywhere. 10 randomly selected snippets per lesson were evaluated, each around 15 seconds.

The lag from recording to viewing was measured before and after each lesson. A timer in front of the VR camera emitted a beep, and it stopped when the beep was heard from the student HMD. In these timing tests, the earphones were removed in favor of the speaker of the phone within the HMD.

Finally, the lag from submission of the student text messages to reception was measured, taking into account the time of the "ping" sent by the macro when dictation was stopped to when the message appeared on the instructor's Poll Everywhere console.

Study Results

Student Experience and Outcome Results

The system allowed for the transfer of knowledge in the case of these two students, as the three pre- and post-tests show. Correct answer rates improved from under 30% on average to over 60%. While "I don't know" constituted the majority of answers in the pre-test, it decreased to under 10% in the post-test.

In addition, the survey results show generally positive attitudes towards the VR system and its contents. In Table 3 below, the survey questions are shown.

Table 4
Survey Questions

No.	Questions and Possible Responses
1	How would you describe the overall ease of use? (5-Very Easy to Use, 4-Somewhat Easy to Use, 3-Can't Say Either Way, 2-Somewhat Difficult to Use, 1-Very Difficult to Use)
2	Would you consider this an interesting learning experience? (5-Very Interesting, 4-Somewhat Interesting, 3-Can't Say Either Way, 2-Somewhat Uninteresting, 1-Very Uninteresting)
3	How interested would you be in using a system like this for learning in the future? (5-Very Interested, 4-Somewhat Interested, 3-Can't Say Either Way, 2-Somewhat Uninterested, 1-Very Uninterested)
4	Do you feel you increased your English knowledge about the situation in the lesson. (5-I Think It Increased a Lot, 4-I think It Increased Somewhat, 3-Can't Say Either Way, 2-I Think It Decreased Somewhat, 1-I think It Decreased a Lot.)
5	How would you rate the audio quality experience when taking the lesson? (5-Very High, 4-Somewhat High, 3-Can't Say Either Way, 2-Somewhat Low, 1-Very Low)
6	How would you rate the video quality experience when taking the lesson? (5-Very High, 4-Somewhat High, 3-Can't Say Either Way, 2-Somewhat Low, 1=Very Low)
7	How would you rate the speech recognition experience when taking the lesson.

	(5-Very High, 4-Somewhat High, 3-Can't Say Either Way, 2=Somewhat Low, 1-Very Low)
8	How would you rate the VR180 format when taking the lesson? (5-Very High, 4-Somewhat High, 3-Can't Say Either Way, 2-Somewhat Low, 1-Very Low)
9	Did today's experience change your interest in using VR for learning? (5-I Think It Increased a Lot, 4-I think It Increased Somewhat, 3-Can't Say Either Way, 2-I Think It Decreased Somewhat, 1-I think It Decreased a Lot.)

Figure 4
Survey results from the two tested students
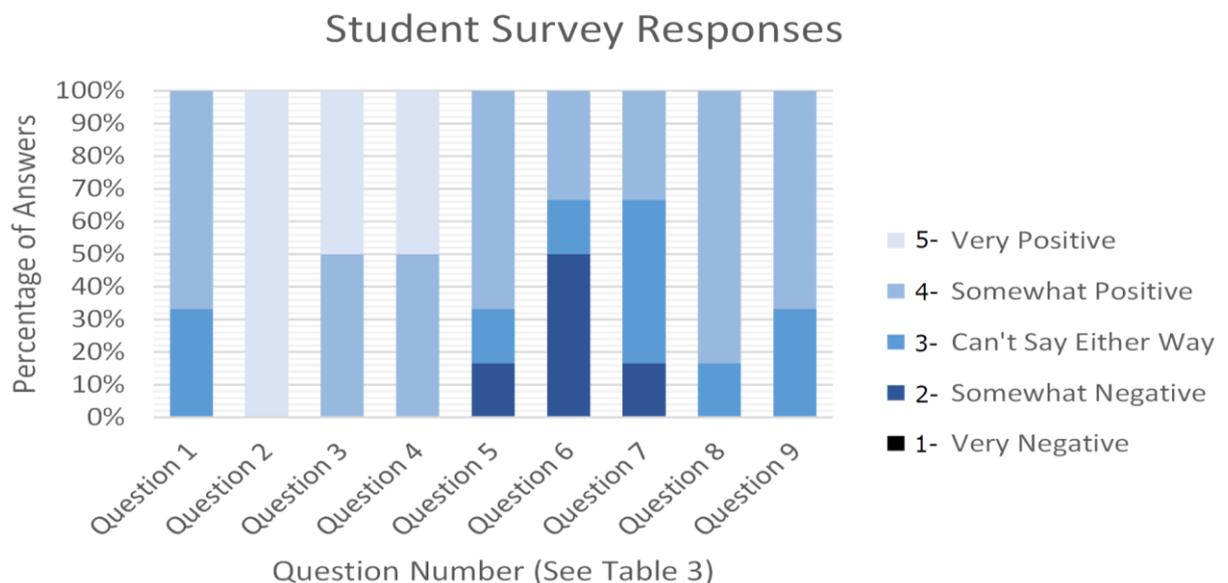
As seen in Figure 4 above, the students generally had positive experiences with the system, but three weaknesses were shown, related to audio and video quality as well as the accuracy of the speech recognition software. This was corroborated by the semi-structured interviews. Each bar represents the average of the Likert scale responses for Questions 1-9 in Table 3, with 5 points for "Very Positive" down to 1 point for "Very Negative." For example, with Question 1, "Very Easy to Use" corresponds to "Very Positive" and "Very Difficult to Use" corresponds to "Very Negative."

Students felt that the video quality was at a somewhat low resolution and that the accuracy of both the macro controls and text-to-speech function could be improved. The issues with macros and text-to-speech were most evident in the first two lessons. Finally, during the semi-structured interviews, both students said they felt more at ease with the privacy offered by not having their likeness displayed during the lesson. One felt that the system protected their privacy boundaries "somewhat more" over a video chat system, and the other felt it was "much more."

From the semi-structured interviews, three main themes were formed. These are "Novel Excitement", "Delay", and "Usefulness", as shown in Table 5 below with their corresponding codes.

Table 5

Themes and Corresponding Codes

Novel Excitement	"Exciting", "First Time", "New"
Delay	"Lag", "Time Spent", "Setup Time",
Usefulness	"Engineering", "Learning", "Concept Clarity"

For "Novel Excitement", the students remained excited about the lesson content throughout the four interview sessions, which adds to the discussion on the idea of the "novelty effect" when it comes to trying new VR experiences (Merchant et al., 2014). Various forms of delay were a major issue with all four sessions, though particularly with the first. Long setup times, lengthy instructions on how to use the system, multi-session periods of calibration for improving text-to-speech accuracy, and the lag between sending communication and the teacher receiving it were all points of contention with both students. The long time between the teacher's broadcasting and students' reception of said broadcasts was nearly a minute, as further explained below, adding to the perception of "Delay." For "Usefulness", both students expounded on the difficulty of explaining complex mechanical movements as required by their field of study, and both felt that the use of 3DVR allowed them to better understand the descriptions of items (with appearance and movement).

System Performance Results

The average lag from recording by the instructor to display on the student's HMD was 46 seconds, with a variation of only ten seconds. This relatively constant lag may prevent the use of direct video or audio communication for students. While YouTube has an option in the livestreaming control panel to set the stream latency (between "Normal", "Low" and "Ultra-Low"), the VR180 App only allows for the "Normal" latency. The lag between the submission of a student text message and its reception by the teacher was on average 3 seconds, with a maximum of 5 seconds. Thus, the average minimum lag of a two-way communication string would be 49 seconds.

The word accuracy rate and correct macro trigger rate are displayed in Figure 5 and Figure 6 below. As shown in Figure 5, there were low accuracy rates for the initial two lessons, particularly for Student 1. After the second lesson, an extensive calibration and training regime of approximately one hour, was conducted with both students. In this, the macro commands were changed to be more recognizable by the software.

Figure 5

The Word Accuracy Rate (WAR%) for Lesson 1 through Lesson 5

Word Accuracy Rate (WAR%)
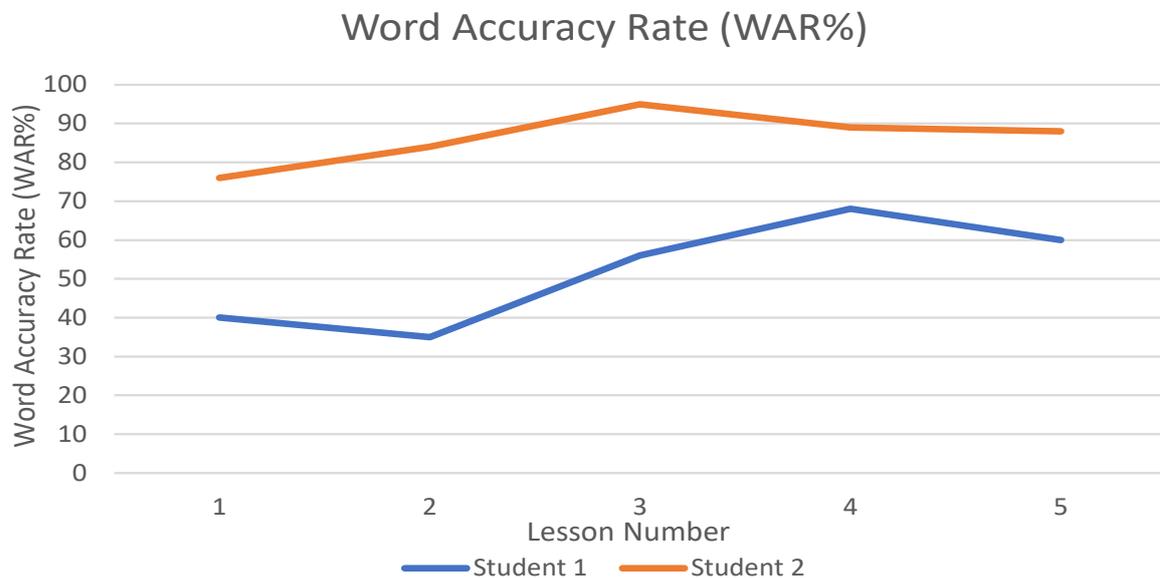
Figure 6

The rate of successful macro activations for Lesson 1 through Lesson 5

Macro Activation Rate
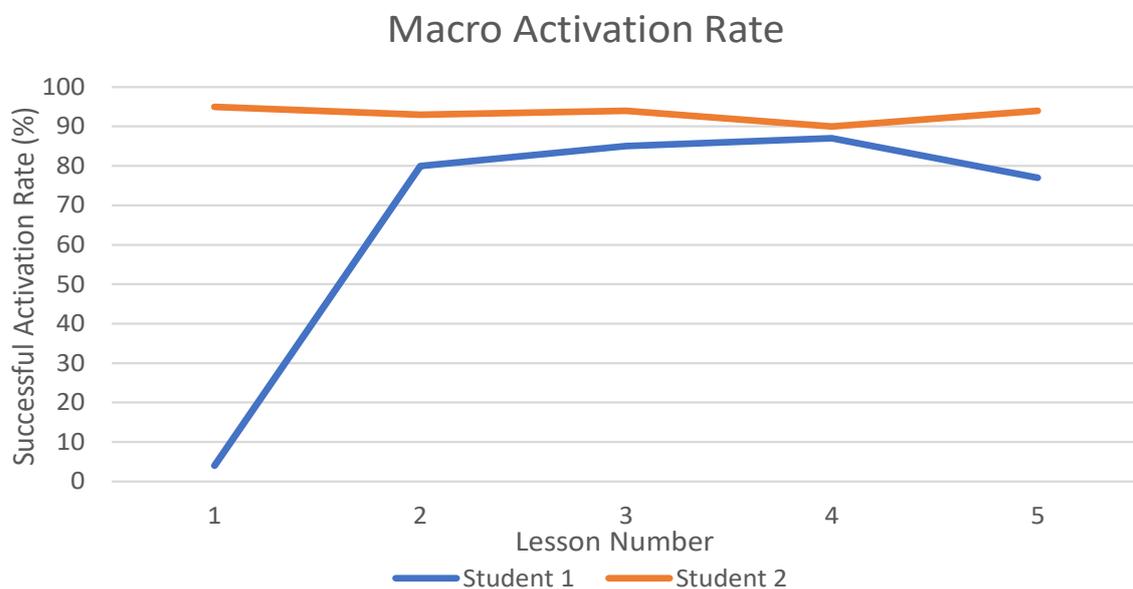
In addition, while the "Voice Macro" program has auto-learning capabilities for speech recognition, a program-specific "training" feature was used several times following the software instructions. For the Cortana speech-to-text program, its "Train Your Computer to Better Understand You" feature was used. In this, the students read multiple lines of text to optimize the speech recognition algorithm. While the relative

effectiveness of auto-learning versus the training features could not be ascertained, there was a clear increase in the accuracy for both macros and speech-to-text.

Through the semi-structured interviews, it was clear that the initial two lessons required great effort to operate the voice recognition controls and speech-to-text functions, and the subsequent lessons were easier.

Discussion and Conclusions

Relevance to Classroom Adoption of Similar Systems

Even relatively inexpensive, easily-available VR setups face great difficulty with mass adoption in the classroom, particularly as a classroom replacement for an entire body of students. Large-scale use of even mid-range VR equipment does not seem to be progressing in higher education beyond the conversion of students' personal devices or class devices into "cardboard-based" HMDs.

There are four main types of HMDs that can fully display VR180 video. The first, and often most basic, are phone-based devices that require a smartphone with gyroscopic sensors as the base, and a holder with the required lenses is attached to the phone. This holder may be very simple, with the framework made of cardboard, or quite complicated, with its own controls and sensors (e.g. Google Daydream, Samsung Gear 360). The second main type is a PC-based headset. These usually require powerful PCs with high-end graphics processing hardware, but they offer fast frame rates and high visual quality (e.g. HTC Vive, Oculus Rift). The fourth main type is a standalone type. This may use a compact processors and operating systems often found in cell phones (e.g. the Oculus Go or Lenovo Mirage Solo). A new category, which this research dubs "hybrid" systems, are multipurpose, with features taken from both standalone and PC-based systems. For example, the Oculus Quest can run relatively resource-non-intensive programs in a standalone mode but can also be connected to a PC for more demanding applications.

In conducting this research, it was apparent that the system created here would not be easily transferrable among VR headset types, and proprietary software locks on some consumer headsets prevented their use with the programs required (for example, the Oculus Rift does not have a usable application that allows for viewing YouTube livestreaming sessions). This would imply that the idea of a "bring-your-own-device" lesson for VR is still far in the future for large-scale use.

VR180 cameras inhabit a similarly wide gamut as headsets, with dozens of manufacturers. They range from very basic cameras with entry-level optics (e.g. Lenovo Mirage, Vuze XR) to studio-oriented systems that use professional-grade sensors and cameras, (e.g. Z CAM K1PRO, Entaniya RIG-3D). In preliminary testing, using two entry level cameras (the examples mentioned earlier), there is a very shallow depth at which there is enough detail to capture written text. However, with proper mounting very close to the instructor and any materials, these issues can be overcome. Livestreaming through a VR camera requires a reliable internet connection with robust bandwidth, especially for the 4K-and-up resolutions that are the default of most VR180 cameras. If there are streaming issues, multiple students may fall out of synchronization as the feeds buffer.

This research shows that a VR system with entry-level devices can successfully be used as a classroom replacement, at least in the medium-term, for a small group of students being taught in parallel to an existing traditional course. The setup and required equipment did not require deeply specialized skills, space requirements, or high costs, which is are some of the great barriers to classroom adoption (Cooper et al., 2019).

The student surveys and tests demonstrated that the two students felt engaged and interested in the material, and they managed to better understand the complex concepts involved. Such interest extended to both the lesson contents and the use of VR. This is reflected in other cases of evaluating student opinions towards VR (Kaplan-Rakowski & Wojdynski, 2018). The time lag, even though over 30 seconds, was also not an issue for the students. However, the reliability of the speech-to-text function and macro controls were initially unacceptable, and extensive recalibrating and improvement were necessary to work. Thus, such a system may in fact not be useful for short-term or single lessons.

Finally, while it was possible to combine the modules for a working system in this case, the underlying proprietary technologies and platforms upon which the modules run are subject to updates and changes. On multiple occasions during the system planning phases, software updates managed to remove the compatibility between modules. These difficulties, along with a steep learning curve for instructors to create VR content, make it apparent that VR has a while to go before becoming a standalone alternative to other distance learning options for CALL.

Relevance to Language Learning Strategies

The use of active learning, inquiry-based learning, and task-based learning seems to be optimal for VR-based systems, especially those that take full advantage of 3-D content or require high levels of immersion. 2D content, particularly non-interactive videos, does not usually allow students to focus their attention on details outside of the already-chosen point of focus, which limits the scope of inquiry-based learning. (Of course, they can still freely examine details within the pre-made video frame.) While this particular system is controlled by text to speech, it could be expanded to students manipulating 3-D objects using controllers paired with an HMD.

As much past research on distance learning using VR focuses on the use of VR as an individual experience, the ability to rapport with peers, as shown in this project, expands the role of students as discussion facilitators and following personalized learning paths. However, it is only until technologies improve, particularly as the latency for livestreaming VR video on common platforms is decreased, that a greater potential for inter-student interaction will occur.

On the instructor side, there will obviously be required changes to teaching styles. The inability to talk naturally with students, partially brought about by the time lag incurred when using livestreaming platforms, may be an issue. Lesson planning may also have to be more tightly structured, as feedback from students may be delayed or lacking due to communications barriers. For large-scale adoption, pre-made learning platforms will have to be used.

Finally, the controls and instructions for many VR platforms are often limited by a low number of supported languages, and pre-existing VR content is overwhelmingly in English and a handful of other languages. This is in contrast to the often-more-

matured user interfaces and language compatibility for legacy distance learning applications. For CALL use to be supported outside of TEFL, an increase in available content is necessary, which may come with time.

Limitations

One weakness in this study is the small number of participants: only two. However, given that the study began as a method to allow a small number of students to simultaneously receive a similar curriculum to that offered in a larger, Zoom-based distance learning class, a small-n study was chosen. Also, given the limitations of equipment and the current health crisis requiring careful logistics and sanitary measures, a large-scale study would have been unnecessarily risky.

Another area for improvement is the focus on a single theme and set of tasks for entirety of the five lessons. For future research, designing entirely different lesson contents for each week in a longer term, perhaps even for an entire course exclusively conducted in VR, may be more useful for evaluating VR as a possible substitute for face-to-face instruction and commonly-used distance learning paradigms, such as 2D video chatrooms or online forums.

Another major limitation is that the area of competency of VR in education is, if not limited to, most applicable to learning tasks requiring visual confirmation and information processing. While the five-week course in this study was entirely made in this way, and students gained skills in describing objects and their functions, this method may not be as useful for courses emphasizing different skills (for example, listening comprehension and reading). Thus, this research proposes that the use of VR streaming video in TEFL is also highly situation-dependent. For example, visual-based learning, including field trips and learning with objects, may be an appropriate use case..

While there are many hurdles to large-scale educational adoption, including initial reliability issues, the targeted use of high-performance, high-presence VR scenarios may be useful for task-based language acquisition, increasing student interest and confidence, and providing alternative immersive learning methods with a high level of student-teacher interaction.

Relation To Current and Future Work

This paper is part of a series aiming to implement recent hardware and software developments in virtual reality technology and understand what they can accomplish in the TEFL field. These advances include multi-tasking through an HMD-based virtual dash, which allows for the integration of a Learning Management System (Urueta & Ogi, 2019), as well as the repurposing of existing non-TEFL VR experiences for TEFL purposes through the use of a customized web portal (Urueta & Ogi, 2020). These two technologies, as well as the currently-explored livestreaming of VR180, have relatively simple learning curves and are the first steps in examining VR as a full substitute for, rather than a supplement to, more traditional platforms. As lecturing, testing, grading, and attendance may all be conducted through VR, this process may be hastened. Rather than a one-off session or a five-week session, the next step may be a full 15-week course spanning a semester and integrating university LMS functionality.

In a future project, the use of this VR180 video system will be expanded to multiple-stream VR180 videos for creating educational materials. In such a system, students would be able to view a learning scenario from multiple angles in 3DVR through the use of two or more synchronized cameras. As an example, a scenario is currently prepared for teaching English note-taking skills in an academic setting, part of a common curriculum requirement at a particular university. In this setting, two VR180 cameras are set up. The first records a lecture from a hypothetical student's point of view (played by an instructor), and the second would be recording a notebook being written in by the hypothetical student. The student viewing the recordings through an HMD would then shift his or her head up and down to switch streams from the lecture to the notepad. In that way, specialized note-taking skills may possibly be imparted through such an experience. This was designed to address three of the common problems with using a single VR camera: slow focusing, lack of detail in the image periphery, and a very short distance at which detail can be seen.

VR holds unique challenges for teachers hoping to integrate it into their distance learning programs, or provide it as an alternative to them, especially so if used on longer timescales and requiring high levels of functionality. As shown in this research, while VR may be best suited for learning specific skills that take advantage of the medium, recent technological developments may also allow for more widespread and longer-term adoption, even if initially for a small subset of students.

References

Adams, W. (2015). Conducting Semi-Structured Interviews. In Newcomer, K., Hatry, H., & Wholey, J. (Eds.), *Handbook of Practical Program Evaluation, Fourth Edition*. Wiley. DOI: 10.1002/9781119171386

Bryan, J., & Karshmer, E. (2013). Assessment in the one-shot session: using pre- and post-tests to measure innovative instructional strategies among first-year students. *College & Research Libraries*, *74*(7), 574-586. https://doi.org/10.5860/crl12-369

Burton, R. (2001). Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers. *Assessment & Evaluation in Higher Education*, *26*(1), 41-50. https://doi.org/10.1080/02602930020022273

Chen, Y., Lasecki, W., Dong, T. (2021) Towards Supporting Programming Education at Scale via Live Streaming. *Proceedings of the ACM on Human-Computer Interaction*, 259. https://doi.org/10.1145/3434168

Cooper, G., Park, H., Nasr, Z., Thong, L.P., & Johnson, R. (2019). Using virtual reality in the classroom: preservice teachers' perceptions of its use as a teaching and learning tool. *Educational Media International*, *56*(1), 1-13. https://doi.org/10.1080/09523987.2019. 1583461

Defanti, T., Acevedo, D., Ainsworth, R., Brown, M., Cutchin, S., Dawe, G, Doerr, K., Johnson, A., Knox, C., Kooima, R., Kuester, F., Leigh, J., Long, L., Otto, P., Petrovic, V., Ponto, K., Prudhomme, A., Rao, R., Renambot, L., Sandin, D., Schulze, J., Smarr, L., Srinivasan, M., Weber, P., & Wickham, G. (2010). The future of the CAVE. *Central European Journal of Engineering* 1, 16-37. https://doi.org/10.2478/s13531-010-0002-5

Drever, E. (2003). Using Semi-Structured Interviews in Small-Scale Research: A Teacher's Guide. Scottish Council For Research In Education.

Ellis, R. (2003). Task-based language learning and teaching. Oxford University Press. 21–37.

Ellis, R. (2009). Task-based language teaching: sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. *Proceedings of the eLearning and Software for Education Conference (eLSE)*, 2015(1).

Feurstein, M.S. (2018). Towards an Integration of 360-degree video in higher education. Workflow, challenges and scenarios. *Proceedings of DeLFI Workshops 2018 Co-located with 16th W-Learning Conference of the German Computer Society, 2250*(1), 1-12. http://ceur-ws.org/Vol-2250/WS_VRAR_paper3.pdf

Fudge, T., & Williams, L. (2020). Zoom in (But fasten your seatbelt). *Proceedings of the International Conference for Media in Education (iCoME) 2020: Diversity Education in ICT Advanced Society*, 195–201.

Guevara, L., Yépez, J., & Guerrero Idrovo, G. (2020). Virtual reality, a telepresence technique applied to distance education: A systematic mapping review. *Proceedings of the 22nd International Conference on Enterprise Information Systems.* DOI: 10.5220/ 0009561805870591

Jensen, L., & Konradsen F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*, 1515-1529. https://doi.org/10.1007/s10639-017-9676-0

Kaplan-Rakowski, R., & Wojdynski, T. (2018). Students' attitudes toward high-immersion virtual reality assisted language learning. *Future-Proof CALL: Language Learning as Exploration and Encounters – Short Papers from EUROCALL 2018*. DOI: 10.14705/rpnet.2018.26.824

Kavanagh, S., Luxton-Reilley, A., Wüensche, B., Plimmer, B. (2016). Creating 360° educational video: a case study. *OzCHI '16: Proceedings of the 28th Australian Conference on Human-Computer Interaction,* 34-39. https://doi.org/10.1145/3010915.3011001

Kozlova, I. (2018). Task-based language learning and learner autonomy in 3D virtual worlds. In Qian, Y. (Ed.), *Virtual reality in education*. IGI Global. DOI: 10.4018/978-1-5225-8179-6.ch032

Kutnick, D., & Joyner, D. (2019). Synchronous at scale: Investigation and implementation of a semi-synchronous online lecture platform. *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*. 1-4. https://doi.org/10.1145/3330430.3333653

Lavin, R. S., & Beaufait, P. A. (2003). Ezboard as a medium for extensive writing in Japanese tertiary EFL classes. *Language Issues*, 9(1), 37–59.

Lee, L. (2016). Autonomous learning through task-based instruction in fully online language courses. *Language Learning & Technology*, 20(2), 81–97. http://llt.msu.edu/issues/ june2016/lee.pdf

Lutes, P. (2018). Engaging and Motivating Students with Active Learning: Project-and Task-Based Learning. *KOTESOL Proceedings 2018.* 20(2), 361-363.

https://koreatesol.org/sites/default/files/pdf_publications/KOTESOL.Proceedings.2018.pdf

Matsuura, K., Kanenishi, K., Morikawa, T. & Yanos, Y. (2004). Flexible participation for collaborative learning in an asynchronous and semi-synchronous context. *Proceedings of EdMedia + Innovate Learning 2004*, 1518–1523.

Merchant, Z., Goetz, T., Cifuentes, L., Keeney-Kennicutt, W., and & Davis T. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, *70*, 29-40. https://doi.org/10.1016/j.compedu.2013.07.033

Radianti, J. Majchrzak, T., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education,* 147. https://doi.org/10.1016/ j.compedu.2019.103778

Welking, F., Wolf, M., Söbke, H., Londong, J. (2019). How to Record 360-degree Videos of Field Trips for Education in Civil Engineering. *Proceedings of DELFI Workshops 2019*. DOI: 10.18420/delfi2019-ws-120

Sherman, W., & Craig, A. (2018). Understanding virtual reality: interface, application, and design. Morgan Kaufman.

Urueta, S., & Ogi, T. (2019). A TEFL virtual reality system for high-presence distance learning. *Advances in Networked-based Information Systems - The 22nd International Conference on Network-Based Information Systems, NBiS 2019*, *1036,* 359–368.

Urueta, S., & Ogi, T. (2020). Web-portal-based repurposing of VR scenarios for TEFL applications. *ICIGP 2020: Proceedings of the 2020 3rd International Conference on Image and Graphics Processing,* 127–133. https://doi.org/ 10.1145/3383812.3383843

Yang, F., & Wang, S. (2014). Students' perception toward personal information and privacy disclosure in e-learning. *The Turkish Online Journal of Educational Technology*, 13(1). https://doi.org/10.1037/t44647-000

Yildirim, G., Yildirim, S., & Dolgunsoz, E. (2019). The effect of VR and traditional videos on learner retention and decision making. *World Journal on Educational Technology: Current Issues*, *11*(1), 21–29. https://doi.org/10.18844/wjet.v11i1.4005