



## OPEN ACCESS

## EDITED BY

Laurissa Tokarchuk,  
Queen Mary University of London,  
United Kingdom

## REVIEWED BY

Mark Billingham,  
University of South Australia, Australia  
Irene Gironacci,  
Agilent Technologies, Germany  
Wilson Luu,  
University of New South  
Wales, Australia

## \*CORRESPONDENCE

Assem Kroma  
assem.kroma@carleton.ca

## SPECIALTY SECTION

This article was submitted to  
Human-Media Interaction,  
a section of the journal  
Frontiers in Computer Science

RECEIVED 26 May 2022

ACCEPTED 05 August 2022

PUBLISHED 09 September 2022

## CITATION

Kroma A, Grinyer K, Scavarelli A,  
Samimi E, Kyian S and Teather RJ  
(2022) The reality of remote extended  
reality research: Practical case studies  
and taxonomy.  
*Front. Comput. Sci.* 4:954038.  
doi: 10.3389/fcomp.2022.954038

## COPYRIGHT

© 2022 Kroma, Grinyer, Scavarelli,  
Samimi, Kyian and Teather. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# The reality of remote extended reality research: Practical case studies and taxonomy

Assem Kroma\*, Kristen Grinyer, Anthony Scavarelli,  
Elaheh Samimi, Stanislav Kyian and Robert J. Teather

School of Information Technology, Carleton University, Ottawa, ON, Canada

Remote user studies—those where the experimenter and participant are not physically located together—offer challenges and opportunities in HCI research in general, and extended reality (XR) research specifically. The COVID-19 pandemic has forced this form of research to overcome a long period of unprecedented circumstances. However, this experience has produced a lot of lessons learned that should be shared. We propose guidelines based on findings from a set of six remote virtual reality studies, by analyzing participants' and researchers' feedback. These studies ranged from one-session types to longitudinal ones and spanned a variety of subjects such as cybersickness, selection tasks, and visual search. In this paper, we offer a big-picture summary of how we conducted these studies, our research design considerations, our findings in these case studies, and what worked well and what did not in different scenarios. Additionally, we propose a taxonomy for devising such studies in a systematic and easy-to-follow manner. We argue that the XR community should move from theoretical proposals and thought pieces to testing and sharing practical data-informed proposals and guidelines.

## KEYWORDS

XR remote study, remote studies, user studies, HCI research, research studies, extended reality, virtual reality, research taxonomy

## 1. Introduction

User studies are the primary tool of evaluating novel interfaces in human-computer interaction (HCI) research (MacKenzie, 2013; Schmidt et al., 2021) and are regularly employed in extended reality (XR) research (LaViola et al., 2017), which broadly encompasses virtual, augmented, and mixed reality (VR, AR, and MR, respectively). User studies are formal experiments with human participants that establish cause and effect relationships between a factor of interest (e.g., interaction techniques) through careful manipulation of experimental factors, or formally, independent variables, while recording and statistically comparing their effect on dependent variables (e.g., performance metrics). Historically, XR researchers typically conduct such studies in a lab. This is often true by necessity, for example, due to specialized, expensive, or non-portable equipment such as CAVEs, head-mounted displays, or haptic devices. Lab-based studies allow researchers to better control extraneous factors that may influence their results, helping avoid confounding variables and improving experimental internal validity

(i.e., “the extent to which an effect observed is due to the test conditions” MacKenzie, 2013). Despite these benefits, lab-based studies are time- and labor-intensive, limit opportunities for parallelization, require involvement from one (and sometimes more) experimenters, and are difficult to deploy to truly random participant pools, which may impact the validity of statistical assumptions. In contrast, data collection methods deployed outside of labs, such as surveys/questionnaires, phone interviews, and field observations, offer some advantages over lab-based studies. Other disciplines, such as social sciences, commonly use these methods (Frippiat and Marquis, 2010; Bhattacharjee, 2012; Brandt, 2012). Of particular recent interest is so-called remote studies; user studies that resemble traditional lab-based studies, but are deployed remotely with little (or sometimes no) experimenter intervention. Remote studies have become more prominent in response to the COVID-19 pandemic (Steed et al., 2016, 2020; Moran, 2020; Moran and Pernice, 2020; Wiberg et al., 2020; Ratcliffe et al., 2021b). HCI research is inherently interdisciplinary (MacKenzie, 2013) and can benefit from surveying how remote studies are employed in related fields (Frippiat and Marquis, 2010; Chaudhuri, 2020). These approaches are already commonly employed in specific sub-areas of HCI, such as mobile usability (Paternò et al., 2007; Burzacca and Paternò, 2013).

Often, user studies are difficult or impossible to deploy remotely due to the nature of the research itself, e.g., studying the performance of specialized custom equipment that is too expensive/difficult/unique to deploy remotely. We instead focus on studies that have no inherent requirement to be conducted in a lab, e.g., primarily studies on software that could be remotely deployed. We investigate the use of remote studies.

In this article, we discuss different types of remote XR studies. We argue that remote user studies are useful not just as a reaction to pandemics, but rather as a regular format for certain study topics that are well-suited to such methods. We support our argument by presenting several case studies: remote VR experiments we have conducted in our lab since the onset of the COVID-19 pandemic. These studies ranged from single session to longitudinal, across a range of topics including cybersickness, target selection, and visual search. Our postmortem includes an analysis of which of these studies were best-suited for lab-based deployment vs. those which were well-suited for remote deployment. We do not propose remote deployment as a “one size fits all” approach, but rather a set of similar approaches to customize based on each study’s nature. We finally present a set of findings and guidelines based on researcher and participant feedback and experience in conducting or participating in these remote XR studies.

The primary contributions of this article include: (i) a taxonomy of remote/lab-based XR studies, presenting options for systematically enumerating remote studies; (ii) research design recommendations based on our practical experience conducting remote XR studies; and (iii) highlighted

areas for future research based on our findings and the proposed taxonomy.

## 2. Related work

### 2.1. The landscape of remote research studies

Research that occurs external to physical research labs have a long history in other fields. Some of them using data collection methods uncommon to XR research, but common in the social sciences (Frippiat and Marquis, 2010; Bhattacharjee, 2012) that have been proven useful to reach populations that are underrepresented in traditional studies (Frippiat and Marquis, 2010; Iacono et al., 2016). These include phone interviews, text messages, surveys (Frippiat and Marquis, 2010; Chaudhuri, 2020), or Skype interviews (Iacono et al., 2016).

HCI researchers are becoming increasingly interested in conducting remote studies (Steed et al., 2020; Schmidt et al., 2021), particularly in web-based (Ma et al., 2018) or mobile application-based (Paternò et al., 2007; Burzacca and Paternò, 2013) formats. This increased significantly with the COVID-19 pandemic and included study paradigms such as co-design (Bertran et al., 2021; Ratcliffe et al., 2021a). Notably, the premiere conference in HCI, the 2020 ACM Conference on Human Factors in Computing Systems (CHI) saw an increase in papers featuring remote studies to reach 15.2%, up from 10.5% in the CHI 2018 conference (Koeman, 2018). In contrast, 42.5% included lab studies, 15.9% included workshops, 14% included interviews, and 12.4% used field studies (Koeman, 2020). This is likely a result of the pandemic; and a quick observation of CHI 2021/2022 papers reveals a larger proportion of remote studies. Such figures raised interest in the feasibility and perceived limitations of conducting such studies in HCI as well as the stigma, if any, around their use (Ratcliffe et al., 2021a). This interest in HCI/XR remote research could be categorized into two main approaches: 1) Considering new forms of research that occur out of the research labs beyond traditional studies (Schmidt et al., 2021); and 2) iteratively building infrastructure to conduct traditional user studies remotely (Steed et al., 2020).

Schmidt et al. (2021) describe several approaches to shift beyond lab-based studies including: (a) re-using previously collected data, which raises questions of availability and standardization; (b) remote testing using applications devised specifically for this purpose, e.g., web-based prototypes; (c) “piggybacking” experiments into remotely deployed applications or webpages rather than physical prototypes; (d) engaging users through remote communication platforms (e.g., Skype, Zoom), although this presents unique challenges, e.g., the digital divide; (e) sending experiment apparatus to participants’ homes, which we discuss later; (f) using analytic and computational evaluation, however, this cannot replace the

human testing; (g) studying phenomena that occur “naturally” online, such as social media usage trends; (h) using VR to simulate studies that would typically not be run in VR; and (i) design the study to target otherwise inaccessible users. Notably, the use of VR to simulate studies introduces special concerns due to several currently unresolved issues with VR itself, such as cybersickness and input issues (Mittelstaedt et al., 2019).

XR studies are particularly difficult to run remotely. They typically require specialized, expensive, and often cumbersome equipment, necessitating complex setup processes. Custom prototypes are frequently unique, and ill-suited for sending to participants due to risk of being lost or broken. Noting these issues, Steed et al. (2020) argue for new approaches to conducting remote user studies for immersive VR/AR. They propose changing attitudes toward research in HCI, building up infrastructure gradually. Initial studies would employ lab personnel and infrastructure; later stages involve forming external participant pools from participants who already have access to required hardware (e.g., commercially-available head-mounted displays), eventually moving on to participants who are provided with equipment through funded hardware distribution. These terms are attached to some ethical and data validity considerations. Consider that crowd-sourcing platforms—one possible option (Ma et al., 2018)—do not offer quality assurance when it comes to participants and the quality of their participation. Additionally, there are ethical questions about participant compensation and recruiting participants from specific geographical areas. These ethical considerations are even more important in the time of the pandemic (Townsend et al., 2020).

These discussions have yielded initiatives such as the XR Distributed Research Network (XRDRN), a platform modeled after classified websites, designed specifically to aid researchers in finding remote study participants during the COVID-19 pandemic (XRDRN, 2020). Other initiatives such as the Virtual Experience Research Accelerator (VERA) aim to become an accelerator supporting remote XR studies as a sustainable form of research in the field (SREAL, 2022). Usage of social communication platforms such as Slack and Discord to facilitate remote XR studies has also increased. These range from a “distributed-VR3DUI” Slack and an “Immersive Learning Research Network” Discord focused completely on XR research, to an “Educators in VR” Discord requiring multiple “research” channels that allow XR researchers to share remote studies without inundating general chats. Other social media, such as Twitter and Facebook, have also become crucial for sharing remote study calls and to connect researchers and participants (Gelinis et al., 2017).

We argue that (1) remote studies are a sustainable method for valid XR research, and not solely a reaction to the pandemic; (2) the community must shift from thought pieces on remote methodologies to empirically verifying their effectiveness; and (3) XR researchers must devise evidence-based and standardized

data collection and reporting methods. This last point is meant to ensure remotely-collected data is both usable and trustworthy. The problem of unifying reported data has been discussed in various contexts, such as cybersickness research (Gilbert et al., 2021). We anticipate that remote studies will face similar challenges to their legitimacy; this must be addressed at an early stage to maintain integrity of research conducted using these methods.

## 2.2. Researcher-participant relationship

### 2.2.1. Researcher as facilitator

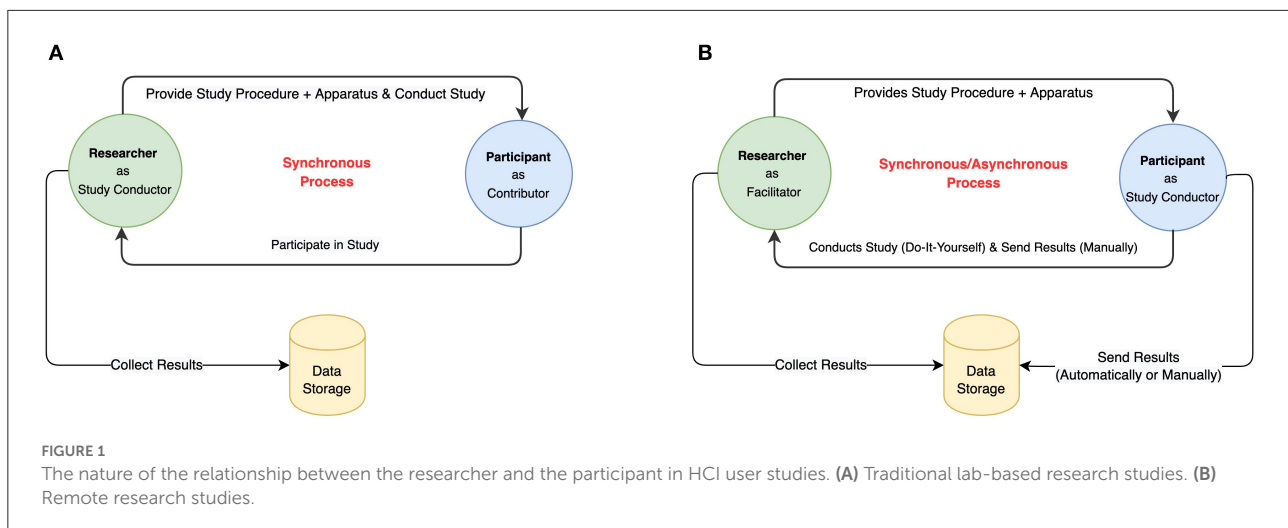
The relationship between the researcher and the participant is unique in remote studies (see Figure 1). In traditional lab-based research studies, the process is synchronous and the researcher is the study conductor. The researcher provides the study procedure and apparatus and conducts the study on participants to collect results to be analyzed (Figure 1A).

In contrast, remote studies can be synchronous or asynchronous, and the researcher acts more as a facilitator than a conductor. The researcher provides the study procedure and apparatus, while the participant conducts the study in a “do-it-yourself” fashion. Data is sent to the researcher either automatically or manually (Figure 1B). Design research workshops employ a similar model where participants conduct research, while the researcher observes and facilitates (Ostafin, 2020; Wala and Maciejewski, 2020).

Remote studies flip the study format in that participants contribute more proactively rather than passively experiencing the experiment procedure. As a side-benefit, remote studies may help spread research culture and empower participants to have more input in future study guidelines, allowing researchers to study both the topic and method simultaneously.

### 2.2.2. Researcher presence

In remote studies, the researcher is either not present or, at best, their presence is mediated *via* video conferencing or avatars in XR environments. Some of our case studies required researcher presence to guide. While there are obvious differences between video/avatar presence vs. in-person, these differences require further study. Having no researcher present may help avoid observer effects influencing participant behavior; researcher presence affects research outcome authenticity (Creswell and Poth, 2017) since participants behave differently in the presence of a researcher. In remote studies, researcher/participant interaction shifts from the traditional synchronous model to become more flexible and either synchronous or asynchronous. Similarly, the lead could shift from the researcher to be either shared or by the participant.



### 3. Case studies

This section presents a postmortem of six case studies in deploying remote VR experiments to identify which features worked well, and which did not. The studies are summarized in Table 1. We present the studies chronologically after grouping by apparatus delivery, which has presented unique challenges and opportunities, and is therefore emphasized. Due to space constraints and scope of this article, we do not reproduce all details of each experiment, instead giving a general overview while emphasizing features of their remote delivery. For complete details of the published studies, we refer the reader to the full references included in Table 1.

#### 3.1. Case study one: Comparing VR platforms within a WebXR learning framework

##### 3.1.1. Overview

We compared the performance differences between three web-based VR platforms supported by the WebXR Device API (W3C, 2019): Desktop, Mobile (tablet), and Head-Mounted Display (HMD). Our study included three components: a Fitts' law (Fitts, 1954; MacKenzie, 1992) selection task extended for 3D scenarios (Teather and Stuerzlinger, 2011) following the ISO 9249-1 standard (ISO, 2000); a "visual search" task involving the placement of one selection target around the individual that they had to find then select to move the target to another position to be found and selected again; and a travel task involving moving through a Virtual Learning Environment (VLE). We assessed each platform's effectiveness through several questionnaires including a self-consciousness scale (Scheier and Carver, 1985), NASA-TLX

(NASA, 2022), and the Intrinsic Motivation Inventory (IMI) (CSDT, 2022).

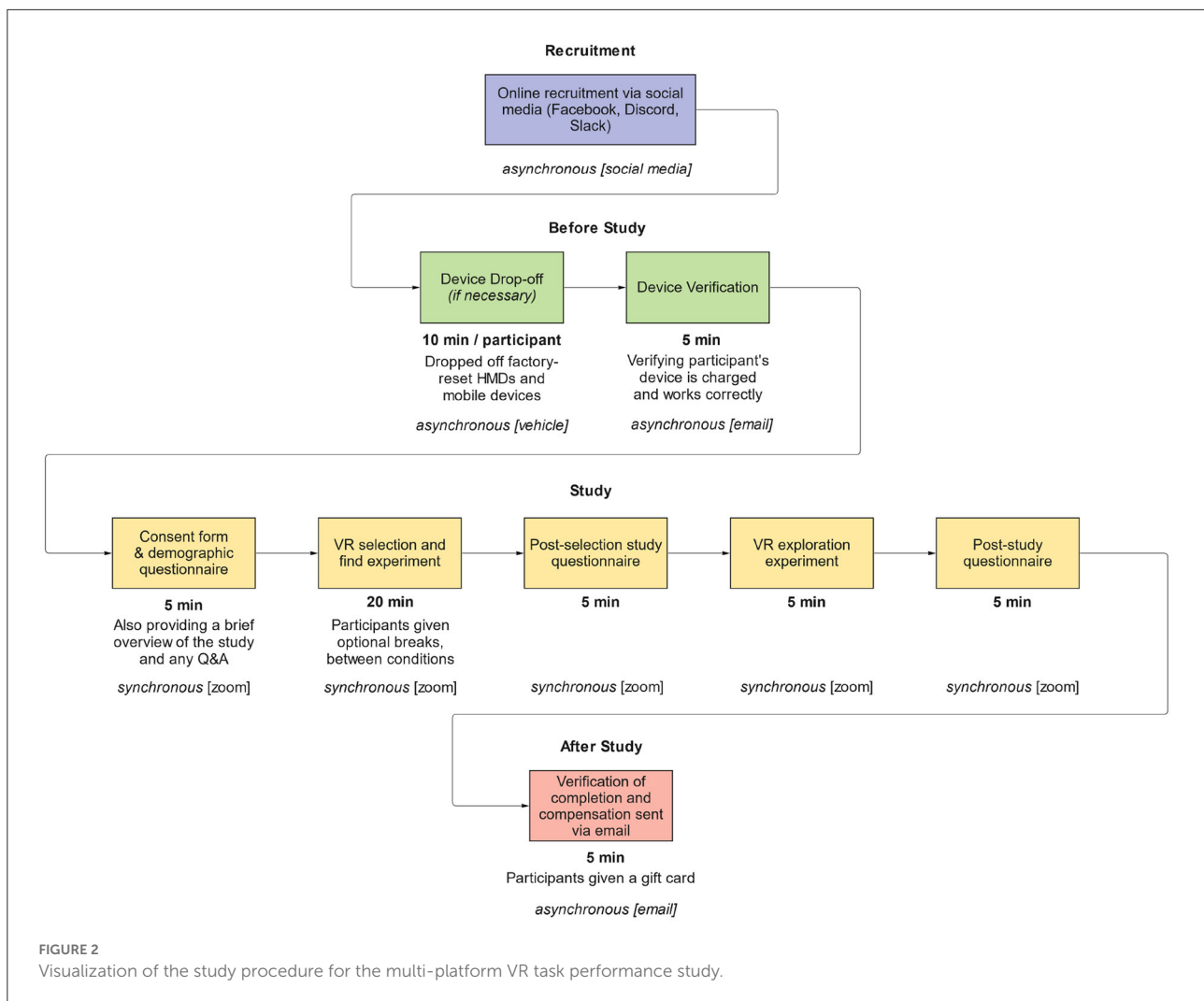
This study was motivated by the observation that many VR studies exclusively compare conditions within a single platform; yet, platform-specific differences such as combinations of display and input devices are likely to become an important factor in determining effectiveness of future multi-platform XR software, where accessibility is of great concern i.e., in classrooms where HMDs may not always be usable. Moreover, as web browsers that support the WebXR specification to allow VR to be viewed by HMD, desktop, and mobile VR platforms become more common it is important to better understand the common interactions and assumptions made in various WebXR developer tools and frameworks.

##### 3.1.2. Methodology

We originally planned to run this study in our lab, employing a within-subjects design but switched to a between-subjects remote study at the onset of the COVID-19 pandemic (WHO, 2022). Finding remote participants with access to all three VR devices (Desktop, Mobile, and HMD) made a within-subjects design impractical due to recruiting difficulty. The study was conducted through the Fall of 2020 and Winter of 2021 where we scheduled synchronous 45 minute video conferencing (Zoom) calls with each participant and using a custom-developed web-based social VR platform called Circles (Scavarelli et al., 2019). The environment included a "Research Room" (see Figure 2) to connect with participants as virtual avatars while they performed the study tasks. The Circles software collected performance data such as selection time and errors, which was made available as a .csv download after

TABLE 1 Comparative summary of the examined XR remote studies.

	Case study	Researcher	Year	Topic/Focus	Duration	Apparatus (SW/HW)	Participants Geo. location	Apparatus delivery	Researcher involvement	Related publication
Contactless drop-off	1	Scavarelli	Fall 2020– Winter 2021	Comparing VR platforms within a WebXR learning framework	One session	WebXR app for use with desktop, mobile/tablet, and HMD (Quest 1)	Local	Contactless drop-off by the researcher	Synchronous	
	2	Samimi	Winter 2021	Selection & repeated refinement	One session	3 apps for use with PC, Samsung Galaxy S8, & Rift CV1	Local	Contactless drop-off by the researcher	Synchronous or asynchronous	Samimi and Teather, 2022a,b
	3	Scavarelli	Spring 2021	WebXR framework for women in trades diversity training	One or two sessions	WebXR app for use with desktop, mobile/tablet, and HMD (Quest 1)	Local	Contactless drop-off by the researcher	Synchronous	
Remote delivery	4	Kroma	Winter 2021 – Summer 2021	Cybersickness/conditioning	Longitudinal	Smartphone app and smartphone for use with google cardboard	Global	Remotely through Amazon	Asynchronous	Kroma et al., 2021
	5	Kyian	Summer 2021	Redirected touch in smartphone-based VR	One session	3 smartphone apps and 2 smartphones for use with VR Shinecon HMD	Global	Remotely through Amazon	Synchronous or asynchronous	Kyian and Teather, 2021
	6	Grinyer	Fall 2021	Field of view & visual search	One session	Smartphone app and smartphone for use with google cardboard	Local	Remotely through Amazon	Synchronous or asynchronous	Grinyer and Teather, 2022



each participant completed their respective selection and visual search experiment.

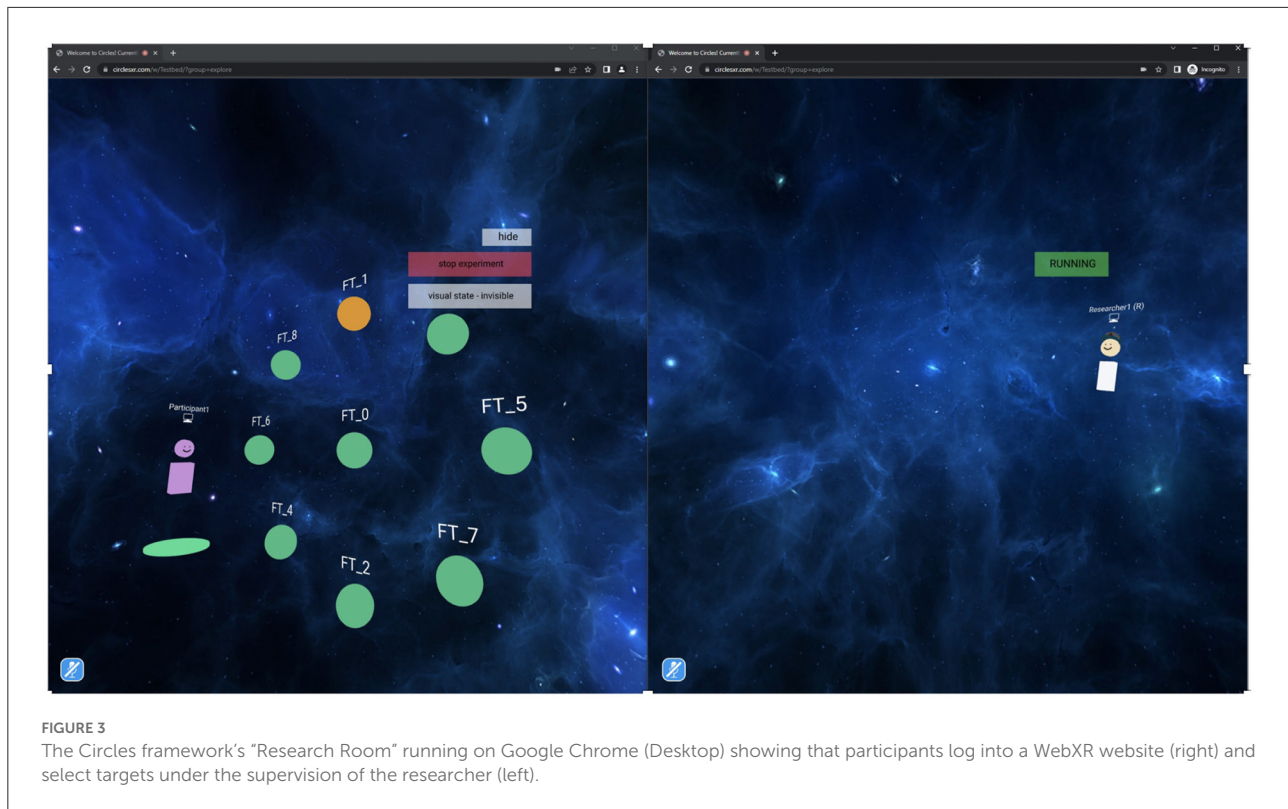
We recruited 45 post-secondary students (18 female, 23 male, 3 non-binary, 1 did not answer) aged 18–44 ( $M = 26.93$  years,  $SD = 7.64$  years), with 15 assigned to each VR platform. We lent out one of 15 Oculus Quest HMDs to all HMD group participants, and for seven of the 15 mobile participants we lent out one of two 10.4" Samsung Galaxy Tab S6 Lite tablets. Each device was sanitized, dropped off, and picked up at each participant's residence after a two week loan period (to provide troubleshooting time). For mobile participants, the other eight participants used their own personal tablets. All desktop participants used their own devices. Most of the personal tablets were 10" Apple iPads and most desktop systems used 21.5" 1920x1080 displays, with a mouse and keyboard. We recorded display resolution, pixel density, and scale so we could calculate target sizes across a variety of tablet and desktop screen sizes.

Participants first completed a consent form and pre-questionnaires *via* SurveyMonkey. Participants then used their respective platform and a Chromium-based browser (Google Chrome on Desktop and Mobile, Oculus Browser for the HMD group) to log into an Amazon-hosted EC2 T2-medium instance of the Circles framework. Circles is a Node.js web application using a MongoDB database with Janus networking running on Ubuntu 20.04. Participants then entered the virtual "Research Room" where they could see the researcher's avatar, who would then direct the participant through the study tasks (see Figure 3).

### 3.1.3. Study discussion

A primary finding of this study is that Desktop provided the best selection performance, likely due to the familiarity of the mouse. The finger-touch reality-based controls (Jacob et al., 2008) of the Mobile were very close, while the HMD "pointing controls" were farther behind. However, visual search





task performance was best with the HMD, perhaps due to the ability to look around a space naturally through head motion. In contrast, the Mobile required holding the device as a "moveable window" with a much smaller field of view. There were no significant differences in social anxiety or social discomfort *via* our questionnaires (Scheier and Carver, 1985), likely due to the online nature of the study, leaving participants less "connected" to the researcher. We note this is a potential limitation of remote studies.

We were surprised to find no differences in motivation by VR platform *via* IMI questionnaire (CSDT, 2022). Immersive VR is often noted as increasing motivation for VR content (Freina and Ott, 2015). This may be a side-effect of delivering the HMDs to participants; in a lab-based study, participants would have limited exposure to the device and thus experience greater novelty. In our case, participants had access to the HMD for roughly a week before the study, and thus could get familiar with the device, offsetting the novelty effect. We also note large variance in all data collected, likely due to the lack of a controlled environment. Users' environments ranged from living rooms to bedrooms to kitchens, typically alone, although sometimes with other people around. Similarly, despite instructing participants on how to fit the HMD, being remote made it difficult to ensure proper fit and correctly adjusted interpupillary distance. It is thus possible that the remote delivery of this experiment has directly impacted the results.

Overall, we feel this study would likely have been better suited to a more controlled lab study to enhance internal validity. Nevertheless, there is some advantage in exploring how these devices are used in more "real world" situations, even if the results are noisier and difficult to parse. Depending on equipment availability, some of this variability might be reduced by providing all equipment to participants, allowing better control across screen size, refresh rate, mouse gain, and screen resolution. Moreover, since the study ran completely web-based (both the VR software platform and all questionnaires), we note there is opportunity to increase participant recruitment by switching to an asynchronous model. This could reduce researcher interference and bias, at the expense of being unable to make sure participants perform the study correctly. A completely asynchronous study could reduce the quality of data collected, but as VR devices become more commonplace and familiar, it may become easier to collect large quantities of participant data using fully asynchronous remote participation.

## 3.2. Case study two: Evaluating the "IMPreSS" selection technique

### 3.2.1. Overview

This study (Samimi and Teather, 2022a) used relatively complex equipment compared to most of the case studies, and

thus we argue is a valuable case study in remote experiment deployment. This experiment evaluated the effectiveness of a novel interaction technique for selecting objects in 3D environments. The technique, IMPReSS (for Improved Multitouch Progressive Refinement Selection Strategy), is a progressive refinement technique for VR object selection that employs a smartphone as a controller, and employs multi-finger selection to more quickly navigate through sub-menus (Samimi and Teather, 2022a). This technique, like other progressive refinement techniques, turns an object selection task into a sequence of smaller/easier selection tasks. Users first select a cluster of objects near the target without high precision. A 2D quad menu that contains all objects selected in the first cluster appears. The user then selects the quadrant containing the target object to refine the selection over several steps until the target object is the only one remaining. Unlike past progressive refinement techniques, we used a smartphone as the controller in this study, and thus could detect how many fingers a user touched the screen with while performing swipe gestures. We used the finger count to indicate which item in a menu should be selected, while swiping in a direction indicates which menu should be selected. This employed the idea of CountMarks (Pollock and Teather, 2020), to improve selection time by reducing the number of menus users drill down into.

### 3.2.2. Methodology

We compared the selection performance of the IMPReSS technique to SQUAD and a multi-touch technique. We recruited 12 participants (4 men, 8 women, aged 26–42 years). Only a third had used VR before, the rest had no prior VR experience.

While originally intended to be an in-person study, we ran this experiment remotely due to the COVID-19 pandemic. After pre-screening participants *via* email, we dropped off the apparatus at their residences. To comply with COVID-19 measures, the equipment delivery was contactless. Participants used their own mouse and keyboard. The apparatus we provided included (i) a PC with an Intel Core i7-7700K CPU at 4.20GHz

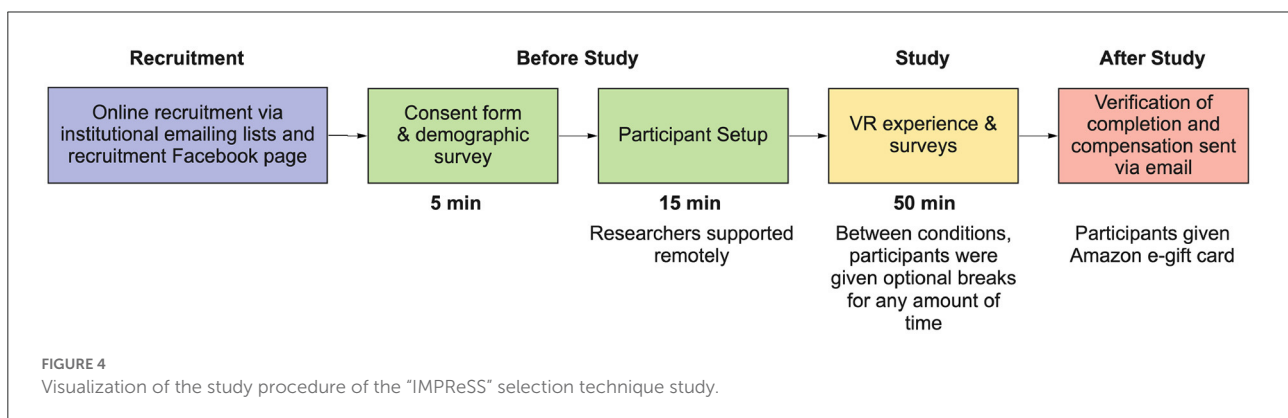
with 32 GB of RAM, (ii) an Oculus Rift CV1, and (iii) a Samsung Galaxy S8 (OS Android 9.0) as the VR controller. We anticipated that setting up this hardware could be challenging for participants, thus we included detailed hardware setup instructions. The experiment software was pre-loaded on the provided PC and Galaxy S8 smartphone. This ready-made setup, while cumbersome, helped ensure that apparatus-related variables were consistent between participants. However, it limited recruitment to a local geographical range.

As shown in Figure 4, after providing informed consent, each participant answered a demographic questionnaire and received a unique ID *via* email. We then delivered equipment as detailed above. For the actual experiment session, participants had to start the software, enter their ID and set the experiment parameters as instructed. The program then started, and the software presented instructions. Pressing any key started the first trial, and put a target object in a random position. After the last trial of each block, the system indicated to the participants that the block was complete. They repeated this process until all trials were finished for all conditions. Overall, the session took ~50 min for the VR part, and ~15 min for setup. A researcher was available to assist participants *via* a video call, thus the study was largely synchronous.

The study employed a 3×3 (Selection technique × density) within-subject design with 8 selection trials for each selection technique/density combination. In total, they completed 3 selection techniques × 3 densities × 8 targets = 72 selections each. The software automatically recorded three dependent variables: search time (s), selection time (s), and error rate (%) to the local computer hard drive. The data was recovered upon receiving back the equipment.

### 3.2.3. Study discussion

Overall, participants were able to select targets faster with the IMPReSS technique than other techniques. Unlike the preceding study, running this experiment remotely did not seem to dramatically affect the results; despite potentially introducing





more variability due to equipment-related challenges, our statistical analysis was still able to detect significant differences between the conditions studied. Notably, this added variability enhanced the experiment's external and ecological validity, while highlighting another challenge in conducting remote experiments, particularly those with extensive apparatus needs.

By delivering the equipment in a contactless manner, the participants had to set up the equipment themselves, which required setup time, consequently limiting the number of trials and densities to keep the experiment duration reasonable. For this reason, we also excluded tutorial/practice sessions. As a result, we did not counterbalance the ordering of one independent variable (distractor density), which likely means that there was a sharp learning curve in early trials, primarily affecting the low-density conditions. This potentially confounding effect is a result of running the experiment remotely.

This also limited our recruitment to participants in a local geographical area. While this is consistent with lab-based experiments, the inability to recruit a global audience, which is possible with other remote methodologies, could be considered a missed opportunity here. This highlights that study-specific priorities dictate the type of remote method employed. For example, if latency is an important factor, then delivering the same equipment to ensure consistency of hardware performance is critical.

Finally, the equipment drop-off process was another limitation of this method, as it slowed data collection. In a conventional lab-based study taking roughly an hour per participant, 8 participants could be scheduled in a given day, facilitating rapid data collection. In contrast, asynchronous remote studies employing Google Cardboard-compatible devices can be run asynchronously, and thus parallelized by deploying to multiple participants at once. In our case, each participant's data took at least 2 days to collect due to sanitizing/quarantining equipment and the scheduling complexity of dropping off and retrieving equipment. Thus, data collection took much longer than either a lab-based study or a remote study using smartphone-based devices.

### 3.3. Case study three: WebXR framework for women in trades diversity training

#### 3.3.1. Overview

Despite efforts to further equality and inclusiveness within STEM education institutions and professions, there is still a lack of female representation within STEM careers (Moss-Racusin et al., 2021), education (Kizilcec and Saltarelli, 2019), and notably the skilled trades (The Ministry of Labour, Training and Skills Development, 2022). We conducted a study that incorporated a VR learning activity into a diversity workshop intended for college faculty. The learning activity highlighted



**FIGURE 5**  
The Circles framework's "Women in Trades" Electrician's School Lab running on Google Chrome (Desktop), from the perspective of a participant. This shows one of the three virtual artifacts—a drill—selected to learn more about microaggressions in learning spaces *via* audio narration, text, and object manipulation.

the challenges women face in the trades within Virtual Learning Environments (VLEs), and our goal of the study was to determine if VR-based learning activities increase motivation (Scavarelli et al., 2020) toward participating in diversity workshops and creating positive behavior changes (Bertrand et al., 2018; Herrera et al., 2018; Shin, 2018) to mitigate gender-based challenges within post-secondary classrooms. Using the Circles WebXR learning framework (Scavarelli et al., 2019), we developed three VLEs. Each VLE showcased challenges women face when (1) growing up, (2) within a trades classroom, and (3) in a trades' workplace (see Figure 5 for an example VLE). These three environments contained several artifacts that allowed participants to explore and learn more about these challenges *via* 3D immersive visuals, text, and narration. The Circles framework allowed all three of these VLEs to be accessed *via* a Chrome-based web-browser *via* Desktop, Mobile, or HMD devices and to visit these virtual spaces with others, in the form of 3D avatars, and communicate *via* gestures and voice.

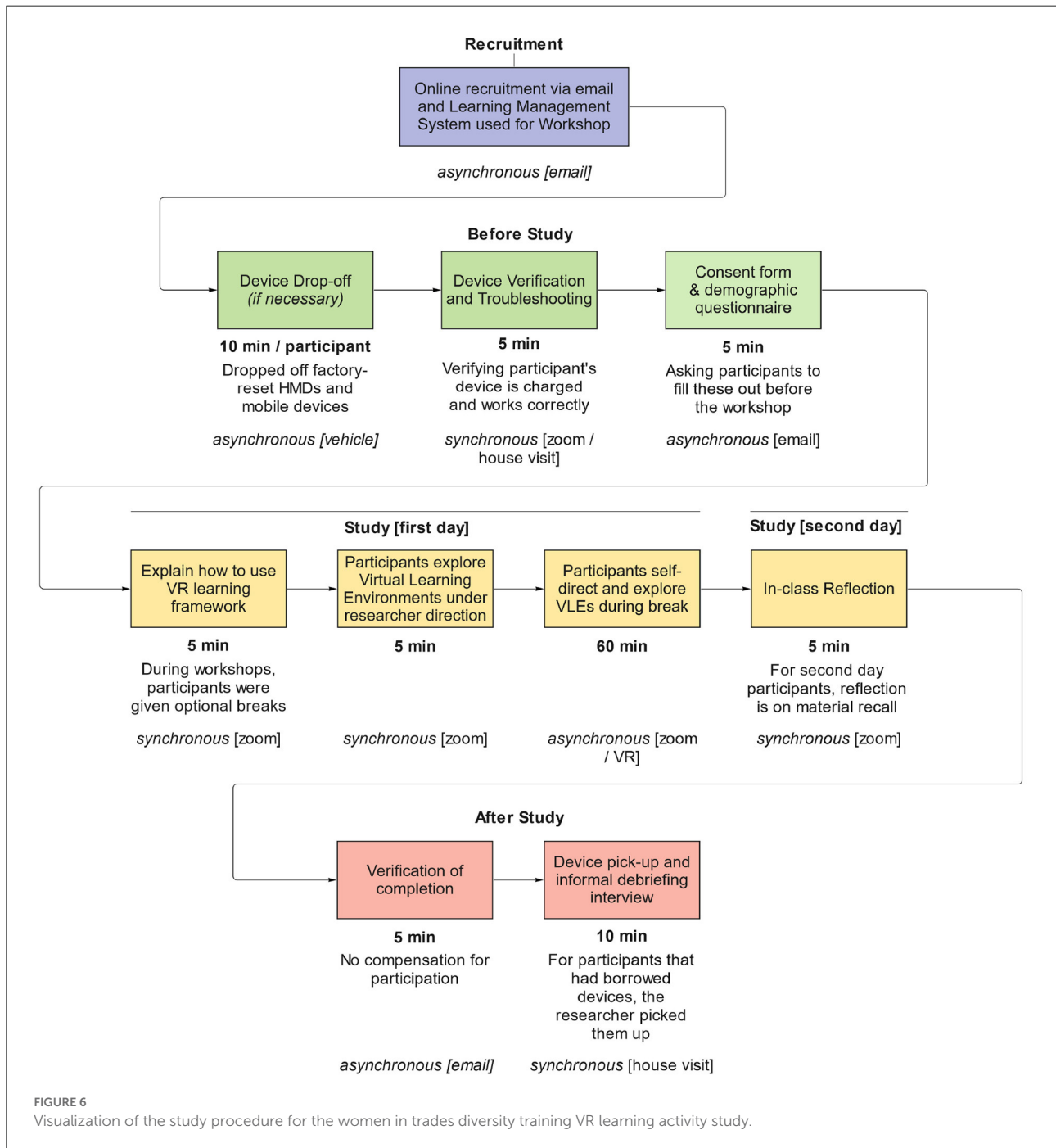
#### 3.3.2. Methodology

This study was "embedded" into two remote gender diversity workshops for faculty run in Spring 2021 at a Canadian post-secondary institution. The COVID-19 pandemic (WHO, 2022) delayed the workshops by a year, and resulted in an online delivery rather than the originally planned in-person model. One workshop ran over 2 days, with each session a week apart. The other workshop took only 1 day, intended for attendees who had participated in a similar workshop previously and wanted a refresher. Both workshops were guided by a facilitator, and used several tools including videos, break-out discussions, and readings to discuss the challenges facing women in the trades/tech. Of the 28 college faculty that attended the workshops, 15 agreed to participate in

this study, though 1 participant later asked to have their data removed leaving 14 study participants (7 female, 6 male, 1 non-binary) between the ages of 18–74 ( $M = 47.21$  years,  $SD = 13.67$  years) in total (out of 28). Participants provided informed consent and basic demographic data through SurveyMonkey. We also collected self-consciousness factors (Scheier and Carver, 1985) in a pre-questionnaire and intrinsic motivation (CSDT, 2022), System Usability Score (Brook, 1996),

and presence (Usoh et al., 2000) in a post-questionnaire (see Figure 6).

Both workshops included a VR learning activity where each participant chose a VR platform with which to access the content. These were run from the participants' homes using our custom WebXR framework, with a backup Zoom call to connect back with the group after the VR experience concluded. Of the 14 study participants, four chose HMD and 10 chose



Desktop. We loaned an Oculus Quest device to each HMD group participant for approximately 2 weeks, and each desktop user used their own personal device. Each Quest was sanitized and dropped off at participants' residences, and retrieved after the study was completed. At the onset of the workshop, we first described what kinds of environments participants would encounter and encouraged exploration and peer discussion on three questions chosen by the workshop facilitator. These questions focused on (1) supporting women considering tech careers, (2) how to make our tech classrooms more inclusive, and (3) how to make women feel more welcome in tech careers. Each participant entered the WebXR experience using a Chrome-based browser from their platform of choice (Desktop, or Oculus Quest HMD). Participants first entered a "campfire" area with three "portals" to other areas. Participants could then re-visit the campfire area to visit any of the other three areas and/or informally discuss their experiences with other participants.

During the workshops we acted as a "participant-observer" (Blandford et al., 2016) to better integrate into the setting and make other participants feel more comfortable. We simultaneously collected observer notes which we then searched for common themes using thematic analysis (Braun and Clarke, 2006) post-study.

### 3.3.3. Study discussion

This study was almost entirely qualitative, so much work was required due to the researcher's role as a participant/observer, collecting data in the form of notes during the VR experience. Device delivery and pickup also necessitated more work than a comparable lab-based study: delivering (and picking up) sanitized devices across the city to participants substantially increased the complexity of this study. Notably, Oculus Quest devices further required a factory reset so that participants could go through the tutorial process and connect their Facebook accounts to the device for use. However, Facebook/Meta's requirement to use a smartphone for setup made participation difficult for some of the participants, likely due to their unfamiliarity with computer technology as more "hands-on" trades faculty. Some did not have a Facebook account or were unable to pair it with the Oculus Quest device, which ultimately meant that two of our HMD participants gave up in frustration and just used their desktop machines to participate. For others, the researcher was able to visit them at their residences using social distancing measures to guide them through the steps.

The fact that the entire workshop was conducted online further increased complexity; it was difficult to control the environment, make sure participants were wearing the headsets correctly, and to observe the relationship between the virtual environments and a shared physical space i.e., a classroom (Miller et al., 2019). However, there were some benefits of conducting the study remotely. For example, we speculate that

we likely recruited a greater number of participants for the workshop and study by allowing remote connections than would be possible in-person. More interestingly, there were some virtual "watercooler" effects (Lin and Kwantes, 2015) observed. For example, participants used the built-in voice communication of Circles to talk with each other informally during the VR trials about the VR experience itself, the virtual artifacts and their relationship to the workshop's goals, and even off-topic discussions such as about the weather or teaching schedules. Some participants even noted "how cool" it was to be able to virtually meet around a campfire and visit the various VLEs. One professor noted immediately that the electrical lab looked just like their own at the College which they had not seen for a few months.

## 3.4. Case study four: Longitudinal study on cybersickness reduction

### 3.4.1. Overview

This longitudinal study (Kroma et al., 2021) was conducted to assess the effectiveness of a common motion sickness conditioning technique, the Puma method (Puma and Puma, 2007; Puma Applied Science, 2017), on cybersickness in VR. Cybersickness commonly occurs in navigating the virtual environment by viewing it through a HMD while in a stationary position due to the lack of physical motion cues. This cybersickness affects presence and performance negatively (Mittelstaedt et al., 2019; Weech et al., 2019). As a result, various techniques to reduce cybersickness have been developed.

The goal of this study was to evaluate the benefits of conditioning techniques as an alternative to visual cybersickness reduction methods [e.g., viewpoint restriction (Fernandes and Feiner, 2016) and viewpoint snapping (Farmani and Teather, 2020)] or habituation approaches that "train" the user to become acclimatized to cybersickness (Howarth and Hodder, 2008). We compared three techniques—habituation through repetitive exposure to VR, the Puma method conditioning exercise, and a placebo (Tai Chi-like movement)—in a cybersickness-inducing navigation task over 10 sessions.

Such a study would be difficult to conduct in a lab under normal circumstances, but would likely be impossible during the pandemic. We chose to conduct this study remotely and use it as a case study in remote user studies due to its unpleasant effects, the large amount of required self-reporting, and the high degree of required participant commitment. We argue that even under normal circumstances, such a study would be easier for participants to conduct at home.

### 3.4.2. Methodology

This study was conducted from Winter through Summer 2021. We recruited fourteen participants (8 female, 6 male, aged

28–40 years), who were recruited globally. As a result, they came from different parts of the world and had different ethnic backgrounds. A demographic questionnaire revealed that the participants had a wide range of experience with VR systems and different frequencies of using VR and playing games on a weekly basis. They self-reported past cybersickness and motion sickness experiences, and completed the motion sickness susceptibility questionnaire. As anticipated, one participant dropped out and another was an outlier, therefore, we used 12 participants in our results. However, we used all 14 participants' qualitative feedback on their experience as a remote study participant.

We divided participants into three groups, see [Figure 7A](#). This study required two of these groups to go through two short (15 min) training/VR sessions each day with a 1–2 h break between sessions, for 10 days as shown in [Figure 7C](#), i.e., 20 sessions in total. The participants had to complete a simulator sickness questionnaire on Google Forms, before and after each session. The sessions included two components: 1) a training session which included video-recorded guidance presenting a cybersickness acclimatization method or a placebo (depending on the group) and 2) a VR session which involved moving through an environment. During the VR experience, participants also responded to a visual survey (10 grades visual Likert scale) to report nausea levels every 2 min. The third “control” group went through one short VR session each day for 10 days as a baseline (control group) to compare the two treatment groups (Puma method and placebo) as seen in [Figure 7B](#). We provided the participants with flowcharts to help them in reviewing the protocols easily whenever needed; these can be seen in [Figures 7B,C](#).

To conduct such a study in a lab would require a tremendous time commitment from each participant, as each would have to come to the lab 20 times. Instead, this study was conducted asynchronously, which allowed the participants to go through the sessions at their convenience, eliminating travel time to and from the lab. We sent a Google Cardboard head-mounted display to each participant, with which they used their own smartphone to display the VR environment. These smartphones were Android-based, and most of them were Samsung phones (ranging from Samsung Galaxy Edge S7 to Samsung Galaxy S21 Ultra). The participants kept the Cardboard as a form of compensation after completing the study. These Cardboards were sent remotely using a direct delivery method through Amazon. The actual software used for the study was sent remotely and downloaded by participants from a password-protected link. The application was developed using Unity and would send the results of the nausea level, the time of the session and basic information directly once the participant hit the “send” button by the end of the session. In case they forgot to do it for one session, the information wouldn't be lost since the log file is accumulative and would be sent during the next session.

### 3.4.3. Study discussion

The remote format of the study does not seem to have affected the study outcome. Preliminary results of this study indicate promising effects ([Kroma et al., 2021](#)). However, we are in the process of conducting further evaluation with a larger participant pool to better generalize our findings, and confirm the effects observed in this study.

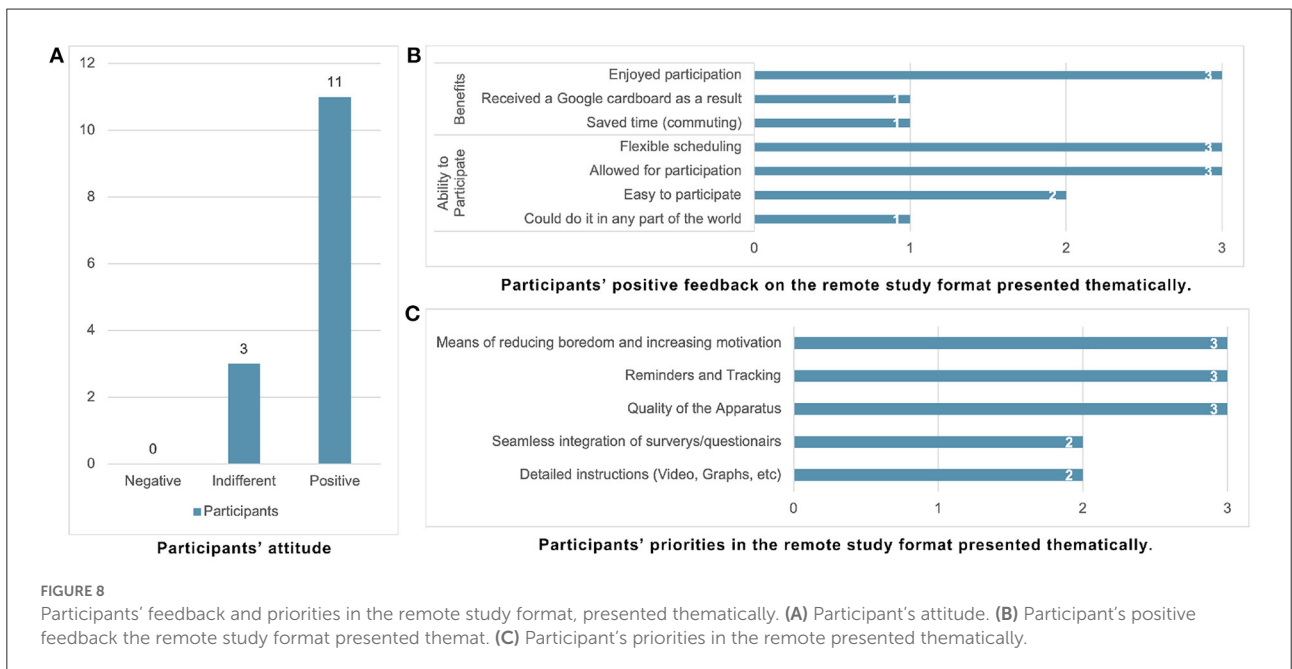
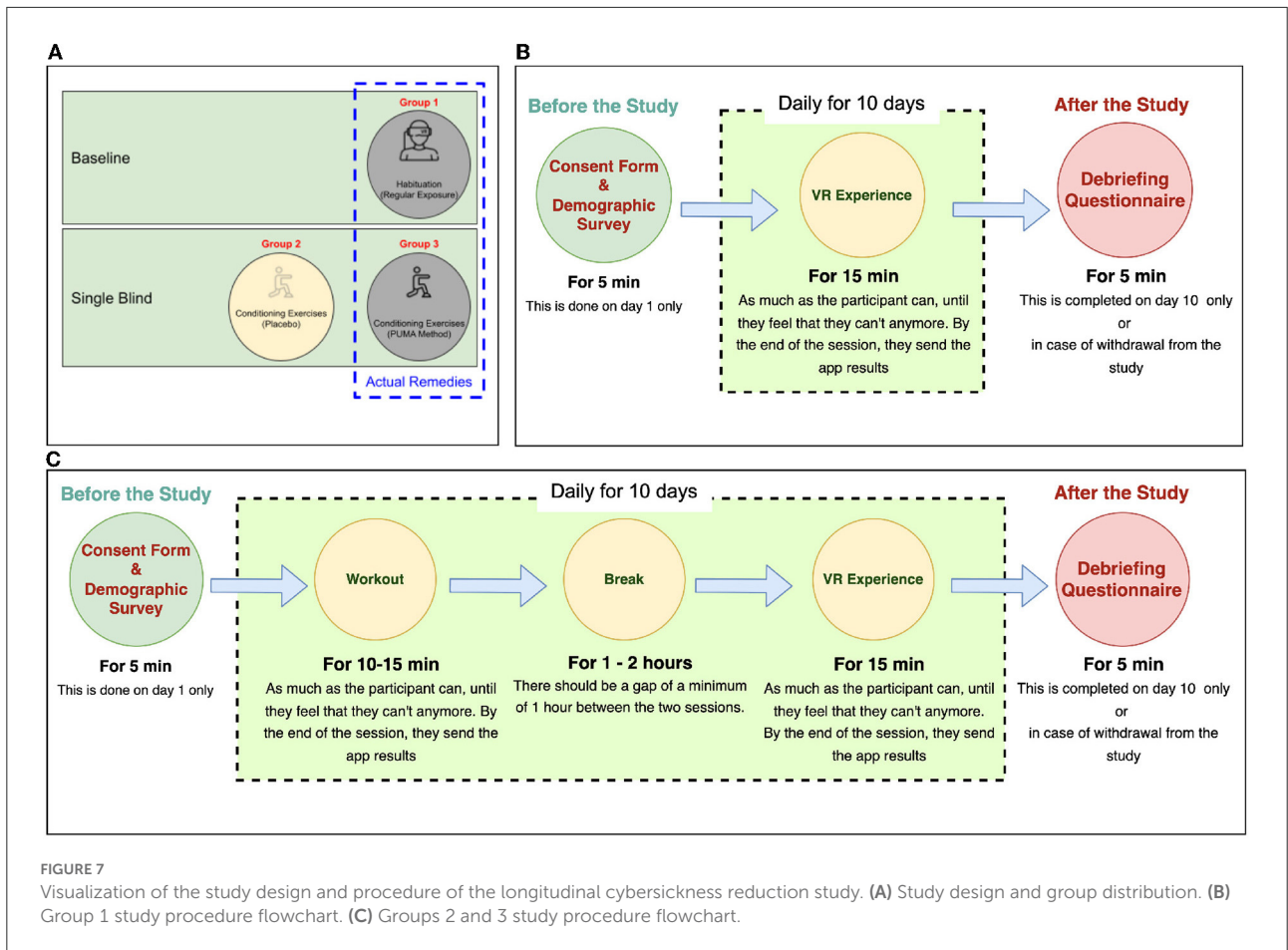
Since this study was conducted as a case study of the feasibility of such a remote format in longitudinal studies, we assessed the participants' attitudes toward the format of the study. Therefore, the participants answered two open-ended questions, and in some cases, the answers were clarified in a follow-up semi-structured interview. Notably, out of 14 participants, only one participant dropped out due to reasons unrelated to the study delivery model. This is consistent with the number of dropouts in cybersickness lab-based studies.

The participants were asked about their thoughts on the remote format of the study. Overall, 11 participants expressed positive feedback about the study, and only three felt indifferent, while no participants gave any negative feedback (see [Figure 8A](#)).

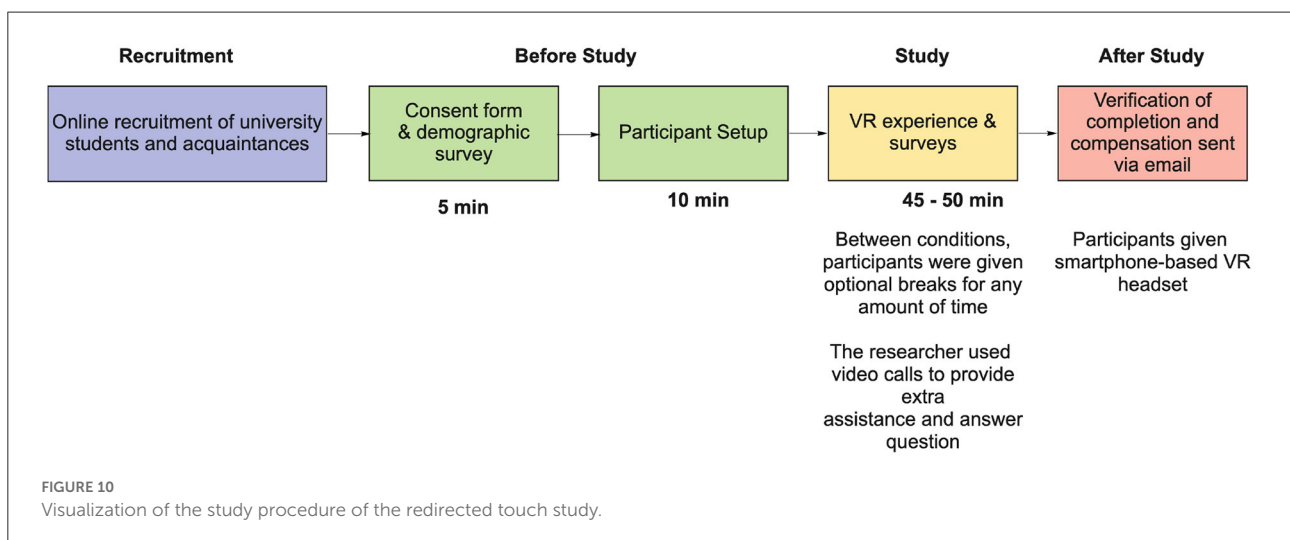
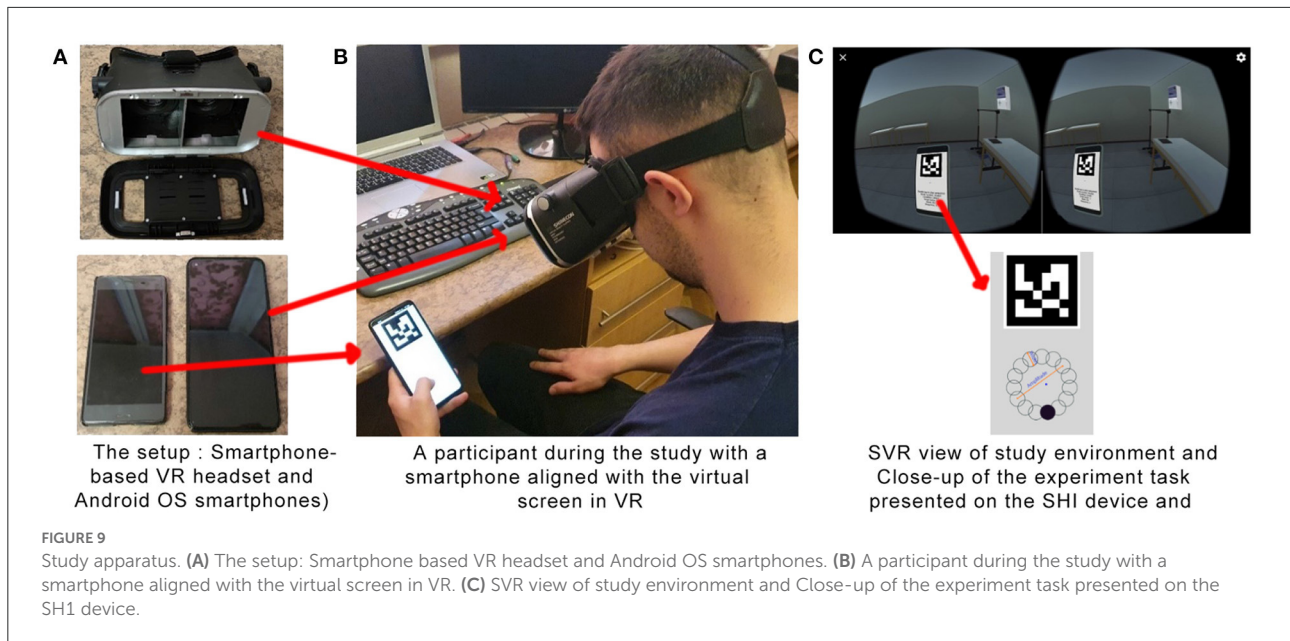
The positive feedback from the 11 participants could be categorized thematically into two main groups: benefits, and ability to participate. Of these, six participants highlighted that the remote format saved them time, while three of these six mentioned that the flexibility helped with their busy schedules. Interestingly, three indicated that they would not participate in such a longitudinal study *at all* had it required going to a lab. This was evident since 91% of the participants conducted their sessions at night. Notably, many participants reported enjoying the study and appreciated receiving clear visual instructions ([Figure 8B](#)).

The participants appreciated the video instructions that guided them through the required activities with the pauses and the counting built-in. They also appreciated the written study manual which included the study flowcharts shown in [Figures 7B,C](#). We provided them only with the chart for their group. They found both of these helpful as they made it easier to understand what was required from them and gave them confidence that they are doing the study properly despite the absence of the researcher. Note that the researcher was reviewing participant progress daily and was in regular contact with participants throughout the study period. It was clear from conversations with participants and their feedback that they took ownership of the study and were enjoying it once they understood what was required.

When asked about what they liked or would like to change in the study, there were three most common priorities ([Figure 8C](#)). These are motivation-related, reminding/tracking-related and apparatus-related. The first two are quite interesting since they are related to both the remote and longitudinal aspects of the study.







### 3.5. Case study five: Redirected touch in smartphone-based VR

#### 3.5.1. Overview

This project (Kyian and Teather, 2021) included two remote user studies to evaluate a method to track a smartphone in VR using a fiducial marker displayed on the screen. Using WebRTC transmission protocol, we captured input on the smartphone touchscreen as well as the screen content, copying them to a virtual representation in VR. To evaluate the effectiveness of our approach under varying degrees of co-location of the control and input spaces, we conducted two Fitts' law experiments. The first compared direct to indirect input (i.e., virtual smartphone co-located with the physical smartphone,

or not). The second experiment assessed the influence of input scaling, i.e., decoupling the virtual cursor from the actual finger position on the smartphone screen so as to provide a larger virtual tactile surface. Both experiments involved sending Google Cardboard-compatible devices remotely to participants, who ran the experiment using their own smartphones.

#### 3.5.2. Methodology

The two experiments were similar methodologically and were both designed and conducted remotely in a contactless manner during the COVID-19 pandemic, in summer 2021. We recruited 12 participants for each experiment. The first experiment included six men and six women (aged 19–49), and

the second study included seven men and five women (aged 18–50). Participants' prior experience with VR varied from “never” to “several times per month”. Eight participants participated in both studies. Participants were recruited globally, from university students and acquaintances *via* email. Participants did not receive compensation but were provided with a Google Cardboard-compatible VR Shinecon (valued at \$15 CAD) that they kept after the study.

Both experiments required substantial apparatus setup, likely the most complex of all of our case studies. To participate in these experiments, participants required two smartphones. One, the SVR device, acted as a VR display and computing device housed in the VR Shinecon device. The second smartphone, the SHI device, was used for hand input (see [Figures 9A,B](#)). We ordered VR Shinecon devices online and shipped directly to participants' homes to avoid direct contact. Since each participant used their own two smartphones, there were substantial differences in the apparatus between participants. We thus recorded the smartphones used for both the SVR and SHI devices for each participant. These ranged from a Redmi Note 7 to Samsung Galaxy S20 Ultra.

The experiment required three applications to be installed on the participants' devices. The first, the FittsStudy application, ran on the SHI device and was based on an existing Unity project FittsLawUnity by [Hansen et al. \(2018\)](#) provided with a BSD-3-Clause license. The software presents a Fitts' law reciprocal selection task for different input devices such as a mouse, eye gaze, or joystick. We modified the code to support finger input, made adjustments to the logs, study configuration settings, and added Android OS support. The app sends the display condition parameters to the RTC app *via* the Android Broadcast method. To track the device during the study, we incorporated a fiducial marker image on the upper part of the screen (see [Figure 9C](#)). Finally, we implemented a script that emailed data logs directly to the experimenter at the end of the study. This allowed for ease of collecting the data and avoiding participants errors, streamlining data collection in this remote study.

The RTC app also ran on the SHI device, and copied the smartphone screen from the SHI device to the virtual smartphone displayed in the SVR application *via* a local Wi-Fi network at 1280×720 pixels resolution. This was sufficient to clearly render the screen in SVR and minimize the latency at the same time. Finally, the VRPhone application ran on the SVR device, and presented a virtual environment seen by the user. It displayed a virtual smartphone, optionally co-located with the physical smartphone, tracked by the SVR smartphone's outward facing camera and the fiducial marker displayed on the SHI screen. Altogether, these applications allowed us to track a smartphone, capture its screen and input events, and present a virtual replica (optionally with touch input scaling/redirection) in VR.

As shown in [Figure 10](#), before starting a study, participants completed consent forms and a demographic questionnaire in

an online format. The instructions included information on how to set up and use the software on smartphones and a description of the study. We used video calls with the participants during the studies to provide extra assistance and answer questions.

Both studies employed the Fitts' law reciprocal selection task. Participants were instructed to select the targets as quickly and as accurately as possible. Regardless if a target was hit or timed out, the next target in the sequence activated. Error trials were not redone but were excluded from the statistical analysis.

The first user study used a within-subjects design with two independent variables: Input Mode (direct, indirect) and Index of Difficulty (ID, ranging from 1.1 to 3.6 bits). The second user study used a within-subjects design with two independent variables: Scale Factor (1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4 scale), and Index of Difficulty (ID, ranging from 1.1 to 3.6 bits). We recorded movement time and throughput. These were recorded and emailed to the researcher automatically, to streamline data collection.

### 3.5.3. Study discussion

The most important result of the study is that the difference in selection throughput between direct and indirect input was both very small and considered statistically equivalent per the result of our equivalence test. This suggests that when using a smartphone in VR, the direct (i.e., real world-like co-located smartphone) condition is not necessary. Overall, our results were generally in line with our hypotheses, and the fact that we were able to reliably detect statistically different performance (and equivalent in the case of equivalence tests) suggests that our data did not present excessive variability.

Notably, we used a relatively small sample size of participants for these two experiments (12 in each, eight in common). There are two reasons for this. First, and perhaps most critically, it was difficult to find many prospective participants willing to participate who also met the hardware requirements (i.e., two available Android-based smartphones). We speculate that this will always be an issue with remote experiments that rely on participants already meeting specialized hardware needs. In our case, the requirement was owning two relatively modern smartphones; studies relying on participants to already possess high-end head-mounted displays (e.g., Oculus Quest 2) would likely face further recruiting difficulty, and perhaps more critically, would likely yield a comparatively homogeneous participant pool of VR-savvy users. This may yield an undesirable shift in results, compromising external validity. Second, we were conscious of the time commitment of conducting these studies, and wanted to keep this relatively low in the event that data collection was compromised, and additional participants were required.

Another complication is related to the apparatus setup. In general, since participants had to set up the software and equipment for the remote study on their own, the setup process

took roughly twice as long as a similar set-up in person. During in-person studies, a major apparatus setup is usually only done once; only minor changes are made for each participant. Unfortunately, this is not the case when each participant uses their own hardware. Thus, we recommend reserving extra time for system setup, especially for participants less comfortable with technology.

Finally, we note a limitation of the hardware itself. We found that due to the heavy graphics requirement, continuous usage of the camera for fiducial marker tracking, and heavy network communication requirements, there was a tendency for the SVR smartphones to overheat during the study. Since the SVR devices are closed in the VR headset (trapping heat in) and the smartphone screen, camera, CPU, and Wi-Fi hotspot are always working, some devices overheated and we had to ask participants to take a rest while the smartphone was cooling down. It is unlikely this issue would occur on a higher-end head-mounted display (e.g., Oculus Quest), but highlights the fact that not all remote methodologies apply equally well across all hardware types. As usual, there is a tradeoff between the convenience of relying on equipment assumed to already be in the participant's possession (a smartphone in this case) and the challenge of dropping off higher-end equipment.

### 3.6. Case study six: Field of view and target search

#### 3.6.1. Overview

This study (Grinyer and Teather, 2022) explored the effects of field of view (FOV), target movement, and number of targets on visual search performance in VR. The goal of this research was to examine dynamic (moving) target acquisition in 3D immersive environments, and the effects of FOV on moving and out-of-view target search as there is relatively little research in both of these areas. We compared visual search tasks in two FOVs under two target movement speeds while varying the visible target count, with targets potentially out of the user's view. We also examined the expected linear relationship between search time and number of items (a relationship well established in 2D environments MacKenzie, 2013) to explore how moving and/or out-of-view targets affected this relationship.

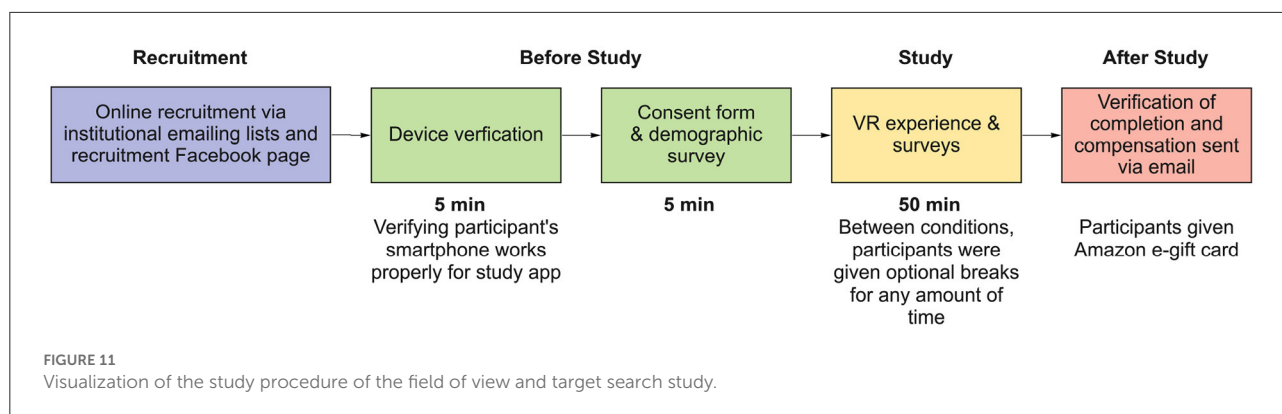
#### 3.6.2. Methodology

A breakdown of our study procedure can be seen in Figure 11. We recruited 25 participants (6 female, 18 male, 1 non-binary). As anticipated, a few participants (two) did not complete the study properly as determined by the recorded data from their sessions. As a result, we only used the data from 20 participants since a participant group divisible by four was needed for counterbalancing. Participants' ages ranged

from 18 to 50 ( $M = 24.75$  years,  $SD = 8.12$  years). We recruited participants by email and through an ad posted on Facebook. Data collection was conducted in the Fall of 2021; during this time, local COVID-19 protocols prevented in-person user studies. Thus, we instead purchased and mailed Google Cardboard VR viewers to each participant. Before sending a Google Cardboard kit to a prospective participant, we first verified that the experiment ran properly on their mobile device. To verify a participant's device, the participant downloaded the experiment application and sent the researchers visual proof the experiment was running normally (e.g., a screenshot of the running application). Once their device was verified, we sent participants a detailed document with step-by-step instructions of the entire experiment process including visuals and links to online surveys. Participants were given multiple options to complete the study with different levels of support from the researchers. Participants could complete the experiment on their own time (i.e., asynchronously), schedule a time with a researcher for instant trouble-shooting support by email, or schedule a virtual meeting with a researcher for on-call support. In all cases, a researcher was always accessible by email, but the first option did not guarantee an immediate response for troubleshooting assistance. All participants but one chose to complete the study asynchronously. The experiment consisted of four conditions with 81 trials (search tasks). In total, participants completed 324 trials requiring they search for and acquire a specific target from a set. We recorded 6,480 trials overall. After completing each condition, the software generated an email containing the recorded data, which was automatically sent to the researcher. This allowed the researcher to verify metrics were recorded properly and participants were completing the experiment honestly and to the best of their ability (i.e., rather than sending a single bulk email at the completion of the experiment). Once a participant completed the experiment, they notified the researcher by email, the researcher verified all parts were completed successfully, and compensation was sent to the participant by email.

#### 3.6.3. Study discussion

We found multiple advantages and disadvantages of employing a remote study. The main advantages were that we could reach a broader pool of participants, as we were less limited by geographic location than a lab-based study. This enhanced the external validity of the experiment since the significant results we found will more likely apply to a broader set of people, situations, and technology setups (MacKenzie, 2013). The largely asynchronous nature of the study allowed participants to complete it at the most convenient time for them. As a side-effect, we saved time in conducting the experiment due to not having to schedule timeslots for each participant. This further saved the participants travel time (and potentially expense). Following analysis of the data, we found



the absence of a researcher did not substantially compromise data integrity; only two out of 25 participants “cheated” and provided inaccurate data. Nevertheless, we cannot be certain each participant performed the study properly as no researcher was present.

On the other hand, conducting a fully remote study presented several disadvantages. With no researcher present during the experiment, it was more difficult to ensure the accuracy of our data; we had to put greater trust into participants to complete the study accurately and to the best of their abilities. As well, participants used different phone models with diverse screen sizes, resolutions, and processing capabilities, potentially causing a varied experience among participants with slightly differing FOVs. These factors likely introduced greater variability in our data than we would expect in a lab-based study, potentially compromising internal validity of our results. Similarly, the variability of devices forced an additional pre-screening step that lengthened our recruitment process. The asynchronous nature of the study resulted in participants completing the study in the late hours of the evening when the researcher was less available to help. This issue could be avoided given stricter scheduling requirements set by the researchers. We also experienced communication issues that would be unlikely in a lab setting. Despite providing detailed instructions, some participants did not properly fill out the parts of the surveys indicating which condition they had just completed. This confusion on the surveys only occurred due to the remote nature of the study; we took extra time to remedy this issue, which necessitated combing through all the submitted surveys to ensure every participant submitted them properly and in order. Moreover, technical issues arose that were unlikely to have occurred in an in-person study. For example, internet maintenance was being performed in one participant's neighborhood while they were performing the study, leading to data loss when data was not sent properly. Lastly, the financial compensation given to participants was greater compared to if the study was conducted in-person due to the increased commitment required by participants.

A key item we took away from completing this study was to minimize the number of steps required of participants when possible. Relying on participants to input any information relating to proper data collection (e.g., entering the condition they most recently completed) should be avoided to minimize errors. We also found it can be beneficial to include additional software-based measures to prevent false data from being recorded. For example, code that detects when participants are skipping trials (i.e., not putting sufficient effort in) would alleviate issues only present during remote studies and help ensure data integrity. Despite these flaws, our experience with conducting remote studies was positive and showed promise for remote studies in the future. We believe with clear communication between researcher and participant and precautions taken to prevent false data from being recorded, remote studies can be as successful as in-person.

### 3.7. Case studies overview

In summary, we had six studies that ranged from one or two sessions to longitudinal (10 sessions). Subject recruitment was online using emails, social media and mailing lists. Apparatus delivery was remotely in 3 of them and through contactless drop-offs for the rest. Data collection was through a mix of self-reporting questionnaires and objective measurements through the deployed XR applications. Participants sent their results remotely without any loss of data as a result of building this function in the applications. Overall, the data collected was satisfactory and tended to be similar to that collected in person.

## 4. Proposed XR remote studies taxonomy

We propose the taxonomy of remote/lab-based XR studies as shown in Figure 12. We originally devised this taxonomy specifically with remote studies in mind, but with flexibility to cover standard lab-based user study options as well. We thus

argue that while this taxonomy helps explore the design space of remote XR studies in particular, it may also be helpful in considering options for XR studies more generally.

The taxonomy is primarily based on the variety of firsthand experiences our group has in conducting remote XR experiments. We further expanded this following brainstorming sessions intended to further consider the design space, and also considering work presented by other authors. We analyzed different attributes that affect choosing the right approach

for remote XR studies and classified their properties and various options.

Our classification (see Figure 12) consists of five dimensions: Study, Participants, Apparatus, Researcher, and Data Collection. Each of these have sub-dimensions and each sub-dimension can have one of the different possible attributes. To this end, it is possible to devise new options for remote XR studies by selecting options from the attributes in the leaf nodes of the taxonomy for each dimension and sub-dimension. We note that although the

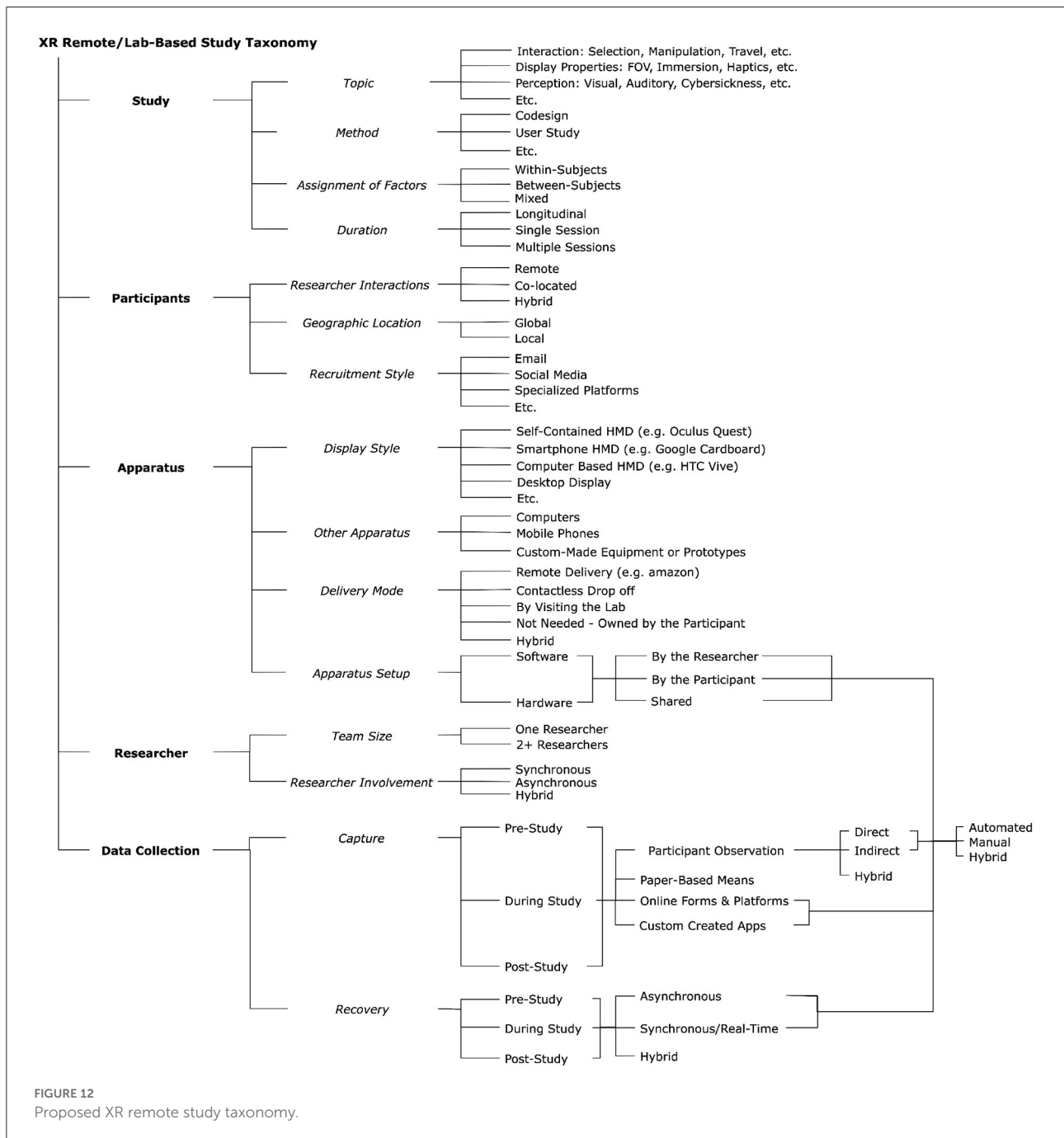


FIGURE 12 Proposed XR remote study taxonomy.



case studies presented above map onto different combinations of attributes offered *via* this taxonomy, they were originally devised in a somewhat *ad hoc* fashion based on the objectives of the individual study and various constraints at the time.

Finally, we note that this taxonomy describes the most universal attributes which we found relevant in literature (MacKenzie, 2013; Steed et al., 2016, 2020; Moran, 2020; Moran and Pernice, 2020; Wiberg et al., 2020; Ratcliffe et al., 2021b; Schmidt et al., 2021; Spittle et al., 2021) and through our implementation of several remote studies detailed above (plus other ongoing and/or unpublished studies). The taxonomy is likely best viewed as a starting point, and we do not claim that it is fully complete. We believe that in the future more attributes will be considered important depending on new scenarios and further applications of remote XR studies.

## 4.1. Study

The Study dimension consists of four sub-dimensions: Topic, Method, Assignment of Factors, and Duration. Topic refers to the broad theme or domain to which the study belongs (e.g., interaction techniques such as selection, manipulation, travel, or other areas such as display properties, or perception.). The Topic is a crucial component in designing any XR remote study because it defines the scope of acceptable limitations, the basic nature of the study, the themes, and so on. For example, latency might be critical in a selection study but not necessarily as critical in XR for education applications. The Method is similarly important since it dictates the framework within which the study is conducted—e.g., a formal experiment, usability study, or co-design, to name a few options. Assignment of Factors indicates how independent variables are assigned to participants, and can be within-subjects, between-subjects, or mixed. We note that this dimension has implications for how a remote XR study is deployed. For example, as described above, some of our case studies would likely have employed a within-subjects design had they been conducted in the lab, but were necessarily deployed remotely as between-subjects designs due to various other constraints (e.g., choice of hardware being prominent among these factors, but time commitment of the participants another likely factor). Finally, the Duration dimension specifies the type and length of study we are using whether it is longitudinal (i.e., many sessions), single session, or over multiple sessions. Multiple sessions means that the study consists of more than a session, but not as lengthy as a typical longitudinal study (e.g., consisting of 10–20 sessions). Multiple sessions would likely be in the range of 2 or 3 sessions.

## 4.2. Participants

The Participants dimension consists of three sub-dimensions relating to the participants of the study: Researcher

Interactions, Participant Geographic Location, and Recruitment Style. Researcher Interactions refers to how participants interact with the researcher while conducting the study. For example, they might be physically co-located with the researcher (i.e., in a lab-based study), remote, or a hybrid approach, for example, being trained locally in a first session, then conducting the remaining study sessions remotely. Geographic Location refers to where the participants originate from. Participants could be local or global, and this has implications for convenience of recruiting, how apparatus is deployed, methodological considerations, breadth of results to different populations (i.e., external/ecological validity), and so on. Finally, there are many options for Recruitment Style, including email, social media, and specialized platforms like Amazon Mechanical Turk. This list is certainly not comprehensive, and other options are also possible.

## 4.3. Apparatus

The Apparatus dimension consists of four sub-dimensions: Display Style, Other Apparatus, Delivery Mode, and Apparatus Setup. Display Style options include Self-Contained HMD (e.g., Oculus Quest), smartphone HMD (e.g., Google Cardboard), Computer-based HMD (e.g., HTC Vive), and desktop displays. Other options are likely possible here, but given the prevalence of HMDs in current XR research, we list a few of these options and note that other display methods (e.g., immersive projective displays) may also be possible, if impractical for remote deployment. Other Apparatus includes any non-display related equipment required of the study, be it computers, mobile phones, or custom-made equipment or prototypes. We have employed a few different options for Delivery Mode, including remote delivery using e-commerce platforms (e.g., Amazon Prime order), drop-off at a participants' doorstep (optionally contactless, e.g., during the pandemic), or by visiting the lab—either to pick up equipment to use it remotely, or in the case of an in-lab study, to physically do the study there.

Finally, we subdivide the Apparatus Setup dimension into Software and Hardware, as depending on the specific study, a researcher may opt to set them up differently. For example, Software can be set up by either the researcher, the participant, or in a hybrid fashion where the participant and researcher share the responsibility of the setup. All of these could be automated (e.g., clicking on a link that takes care of everything), manual (e.g., having to download and/or install software on a system) or a hybrid option somewhere in-between (e.g., where some software installs automatically, other software requires manual intervention). The Hardware setup can similarly be done by the researcher, the participant, or hybrid. While the Hardware and Software Setup may use the same set of options, we note that this is not required, depending on the circumstances and study design related considerations. However, intuitive

and seamless XR applications are crucial for the success of remote studies.

#### 4.4. Researcher

The Researcher dimension includes two sub-dimensions: Team Size, and Researcher Involvement. Both of these have implications for how a remote study might be deployed. For example, Team Size could be one researcher or two or more researchers. Team Size is another factor to consider because parallelization is likely easier with more than one researcher. Researcher Involvement could be either synchronous by attending the actual session, asynchronous by giving advice and reviewing the results after or before the session, hybrid i.e., a mix of synchronous/asynchronous, or alternatively either option being available at the participants' discretion. The latter was the choice in the studies by [Kyian and Teather \(2021\)](#), [Grinyer and Teather \(2022\)](#), and [Samimi and Teather \(2022a\)](#).

#### 4.5. Data collection

Data Collection has two sub-dimensions: Capture and Recovery. Capture refers to the means of collecting data and Recovery refers to the means of receiving this captured data by the researcher. These two each have three sub-dimensions: Pre-Study, During Study, and Post-Study. Each refers to data collected at different stages in the study (e.g., pre- and post-questionnaires, or data collected by software or researcher during the actual study task). For the Capture dimension, each of these sub-dimensions has a similar set of possible options: data could be collected using paper-based means (e.g., forms, notes, sketches), participant observation (e.g., video, live streaming), online forms and platforms (e.g., Google Forms, Qualtrics), or using custom-developed applications (including instrumenting experiment software to automatically collect data). The latter two options (online forms, custom apps) can be automated, require manual intervention, or an in-between hybrid option to collect data. Participant Observation can be direct (e.g., in real-time in-person or through live streaming), or indirect (e.g., through viewing video recordings post-session). These could be automated (e.g., auto-recorded by the application), manual (e.g., recorded by the participant), or an in-between hybrid option.

Similarly, the Recovery dimension includes three sub-dimensions: pre-, during, and post-study. These include similar options that consist of asynchronous recovery (e.g., data is collected with a time gap or at a specified time), synchronous (e.g., data is sent in real time), or a hybrid option in which some of the data is sent in real time while the rest is sent after a time gap). This recovery could be automated (e.g., sent without the

participant's involvement), manual (e.g., requires participant's action), or hybrid (e.g., the participant clicks a button to start an automated process or some of the data is sent automatically while the rest is manual).

## 5. Discussion

### 5.1. Issues with remote studies

While remote studies offer some benefits, they also present some issues that should be addressed. We list and discuss these below. The prevalence and impact of each of these issues is highly dependent on the attributes of the remote study itself. For example, experiments with synchronous researcher involvement and manual data collection may be more forgiving with respect to technical considerations than, for example, a fully asynchronous and automated data collection study, which has much higher requirements on bug-free and "perfect" apparatus implementation.

#### 5.1.1. Technical considerations

There are many technical considerations when conducting remote studies, such as: (1) producing an almost bug-free and potentially totally automated/participant-directed apparatus, which is time-consuming and requires many iterations and extensive testing; (2) protecting against data breaches and security and how to handle related settings on different third-party platforms; (3) accounting for equipment discrepancies if it is not provided to participants, such as phone memory, screen size, quality (all of which increase the likelihood of confounding variables in the experiment) and the possibility of damaging participants' equipment and related liability; (4) understanding operating system-related limitations which could limit the study to one type of headset/phone, limiting the participant pool, and all inconsistency/labor related issues; (5) ensuring headset quality—for example, there is inconsistency in Google Cardboard manufacturing and minor differences between products; and (6) purchasing and delivering equipment and all related logistics, such as availability, prices, delivery dates, returns, and refunds.

#### 5.1.2. Participant-related

There are several issues relating to participants. These include (1) reminding the participants to conduct the sessions in longitudinal studies that take place over multiple sessions; (2) motivating them to conduct the studies; (3) counting on their commitment and integrity in reporting, especially in asynchronous data collection situations; (4) helping them go through their learning curves, which makes the focus shift from the actual study to the learning process; and (5) their technical skills, which means that recruitment leans toward

more tech-savvy participants. This is similar to a common issue in university research labs which predominantly recruit undergraduate students. Reaching the right participants is a tricky task. These participants should not be only suitable, diverse, and representative of the general population, they should be interested and committed to completing the study with integrity. Therefore, initiatives like VERA (SREAL, 2022) and XRDRN (XRDRN, 2020) are great steps in the right direction.

### 5.1.3. Research-related

More generally, there are also research-related issues, including: (1) ensuring the quality of the data, monitoring errors, and the inability to correct mistakes (which could yield unusable data) promptly in asynchronous studies; and (2) ensuring ethical applications and obtaining research ethics board approvals. Remote deployment introduces more possibilities of confounding variables in an experiment, e.g., through different hardware configurations, and different distractions that each participant may be subject to. In contrast, the principle advantage of in-lab studies is the ability to control the environment, thus ensuring internal validity. With a sufficiently large participant pool, these external influences “average out” and likely do not matter as much, but dramatically improve external validity.

### 5.1.4. Logistics

Finally, logistical issues should not be ignored, as they can have a dramatic impact on the quality of data collected and can impose certain choices in the taxonomy presented above. These can include (1) budgeting and cost variances between participants due to currency rates, geographic location, etc.; (2) equipment availability; and (3) shipments and apparatus delivery. For example, delivering equipment to participants on the other side of the globe may be infeasible and impose a different type of apparatus than that originally envisioned if global participation is deemed necessary.

## 5.2. Benefits of remote studies

Remote studies offer many benefits as well. Based on our observations in these six studies, we argue that they could give participants more of a sense of empowerment and ownership compared to lab-based studies. However, some participants might not feel the same or confident in remote XR studies, especially users with limited experience with XR devices and applications. Additionally, it allows for operating on a global scale and having a potentially more diverse and inclusive participant pool, which was noted by participants in some of our case studies. Notably, our longitudinal

cybersickness study included participants from three continents and four ethnic groups. Deployed on a large scale, such results have the potential to be far more representative of broader populations than a typical lab-based study thereby improving external validity. Furthermore, remote deployment helps facilitate conducting longer longitudinal studies and automating data collection. Since participants participate from the comfort of their homes, this may make them more focused on the study rather than being distracted by the researcher; this reduces the aforementioned researcher presence effects on the study. It allows study conditions to be more realistic (i.e., improving ecological validity), especially for longitudinal studies, enhancing experimental external validity (MacKenzie, 2013). Furthermore, remote studies allow for recruiting larger numbers of participants. For instance, CHI 2020 statistics showed that remote studies are typically bigger by comparing participant numbers. The largest remote study was conducted by Facebook, who recruited 49,943 participants *via* surveys and Amazon Mechanical Turk (Koeman, 2020).

## 5.3. Recommendations and guidelines

Reflecting on participant feedback, our observations, and the highlighted issues/opportunities in remote user studies, it is important to consider the following points when designing remote user studies.

### 5.3.1. Seamless study design and automation

Participants are not researchers, and do not have the same training to run experiments. It is thus critical to design experiment pipelines that are easy to follow and flexible to reduce the burden on participants. In our experience, it is important to automate reminders, trackers, and data collection while ensuring privacy and data security. This is especially true in the case of longitudinal studies. Procedural instructions should be easy to follow through clear visuals, flowcharts, and instructions both in text and video formats to ensure clarity and the satisfaction of different audience preferences. Furthermore, it is helpful to think on a global scale and ethnic representation in the participants' pool.

### 5.3.2. Apparatus design

The apparatus has to be “bullet-proof”—hardware should work well and software should be virtually bug-free. Extended reality applications that work independent of head-mounted displays would be ideal; this can be supported through the use of technologies such as WebXR. Such technologies reduce labor and could allow for server-based data collection, which opens the door for recruiting participants using crowdsourcing platforms like the Amazon Mechanical Turk (Ma et al., 2018).

Furthermore, pushing for updates directly to the application without asking the participants to reinstall newer versions could be useful to solve bugs and apparatus issues without participants having to sort out the details. Finally, this also requires ensuring the security of the application, data communication, and storage servers.

### 5.3.3. New opportunities

Remote studies open the door to many interesting opportunities. For example, consider the potential use of Artificial Intelligence (AI) to take on the role of a research assistant. This includes conducting tasks that are clear, from an algorithmic point of view, and allow for more involvement with the participants, which supports quality assurance in large-scale studies. Another opportunity is creating support communities similar to “Unity Pulse,” which is an online feedback community (Unity Technologies, 2021). This improves participant motivation as they can claim volunteering hours and be part of a community. In such a setup, incentives transform beyond the traditional monetary, gift card or artifact. Amazon has opened the door for crowd-sourcing of participants through Amazon Mechanical Turk (Amazon Mechanical Turk, 2018), which has great potential if we can overcome platform-specific limitations and ethical/data-related concerns.

### 5.3.4. Ethical considerations and lessons learned

Another priority is to address ethical issues specifically related to remote studies (Chaudhuri, 2020; Steed et al., 2020). These take different forms, and while general considerations could be addressed, each study has specific ethical issues which are probable, especially with our limited experience with these remote studies.

On the other hand, there have been many remote lab experiments in the paradigm of distance learning (Nickerson et al., 2007; Cooper and Ferreira, 2009; Ma et al., 2018), in addition to remote studies in other research disciplines; these experiences could be a good starting point to learn from.

## 6. Conclusion and future directions

In this article, we summarized six case studies of conducting remote XR studies during the COVID-19 pandemic. Based on our experiences in these studies and the current body of work in this area, we proposed a taxonomy that helps in designing and conducting remote XR studies. We do not claim that the proposed taxonomy is complete, but rather a foundation to providing an easy-to-follow systematic

approach to designing remote XR studies. Additionally, we offered a set of recommendations and guidelines based on these findings to be considered when conducting XR remote studies.

Employing remote studies in XR research presents new opportunities, and should not be merely a reaction to a pandemic. It opens the door to larger participant pools, facilitates easier application of longitudinal studies, and democratizes the technology in the process. Our case studies highlight several opportunities that remote study brings to XR research, similar to those found in the social sciences (Fripiat and Marquis, 2010; Bhattacharjee, 2012). The COVID-19 pandemic has forced XR researchers to adapt and investigate remote study methods, and further examination of our remote experiments has helped to realize that beyond the pandemic, there is a significant opportunity in using remote study in future XR experiments.

Understanding these studies’ benefits and limitations allows for proper employment and the continuation of research studies in pandemics, by building a body of communal knowledge and infrastructure for stronger remote XR studies. Standardizing remote studies in HCI as a whole and extended reality specifically is a community effort that should be built over a period of time by sharing findings and best practices until it reaches a level of maturity that allows for more quality research data collection and findings. It is important to start this process that could benefit the community in many different ways. This allows us to shift from reacting to pandemics to employing new formats of research studies that have shown effectiveness in other fields and a great promise in HCI.

## Data availability statement

The datasets presented in this article are not readily available because of REB required restrictions. Requests to access the datasets should be directed to RT, [rob.teather@carleton.ca](mailto:rob.teather@carleton.ca).

## Ethics statement

The studies involving human participants were reviewed and approved by Carleton University-Research Ethics Board. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.



## Funding

This study was supported by NSERC Discovery Grant RGPIN-2016-06265 and Canada Foundation for Innovation.

## Acknowledgments

The authors gratefully acknowledge the support of the School of Information Technology at Carleton University and the generous research infrastructure and personnel funding by CFI and NSERC. This article would not be possible without the valuable feedback of our lab colleagues, the study participants, and the reviewers.

## References

- Amazon Mechanical Turk (2018). *What is Amazon Mechanical Turk?—Amazon Mechanical Turk*. Amazon Mechanical Turk.
- Bertran, F. A., Pometko, A., Gupta, M., Wilcox, L., Banerjee, R., and Isbister, K. (2021). “Designerly tele-experiences: a new approach to remote yet still situated co-design,” in *ACM Transactions on Computer-Human Interaction* (New York, NY: Association for Computing Machinery). doi: 10.1145/3506698
- Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., and Zenasni, F. (2018). Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Front. Rob. AI* 5, 1–18. doi: 10.3389/frobt.2018.00026
- Bhattacharjee, A. (2012). *Social Science Research: Principles, Methods, and Practices*. Tampa, FL: Global Text Project.
- Blandford, A., Furniss, D., and Makri, S. (2016). Qualitative HCI research: going behind the scenes. *Synthesis Lectures Hum. Centered Inform.* 9, 115. doi: 10.1007/978-3-031-02217-3
- Brandt, C. (2012). *Taking Research Beyond the Lab*. Calgary, AB: Mount Royal University.
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa
- Brook, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluat. Ind.* 189, 4–7.
- Burzacca, P., and Paternò, F. (2013). “Remote usability evaluation of mobile web applications,” in *Human-Computer Interaction. Human-Centred Design Approaches, Methods, Tools, and Environments*, ed M. Kurosu (Berlin; Heidelberg: Springer Berlin Heidelberg), 241–248.
- Chaudhuri, T. (2020). *Collecting Data During COVID-19: How do we Address Data Quality?* Medium.
- Cooper, M., and Ferreira, J. M. (2009). Remote laboratories extending access to science and engineering curricular. *IEEE Trans. Learn. Technol.* 2, 342–353. doi: 10.1109/TLT.2009.43
- Creswell, J. W., and Poth, C. N. (2017). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Los Angeles, CA: Sage.
- CSDT (2022). Intrinsic Motivation Inventory (IMI). Available online at: <https://selfdeterminationtheory.org/intrinsicmotivation-inventory/> (accessed September 12, 2019).
- Farmani, Y., and Teather, R. J. (2020). Evaluating discrete viewpoint control to reduce cybersickness in virtual reality. *Virtual Real.* 24, 645–664. doi: 10.1007/s10055-020-00425-x
- Fernandes, A. S., and Feiner, S. K. (2016). “Combating VR sickness through subtle dynamic field-of-view modification,” in *2016 IEEE Symposium on 3D User Interfaces, 3DUI 2016-Proceedings* (Greenville, SC: Institute of Electrical and Electronics Engineers Inc.), 201–210.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.* 47, 381–391. doi: 10.1037/h0055392
- Freina, L., and Ott, M. (2015). “A literature review on immersive virtual reality in education: state of the art and perspectives,” in *Proceedings of eLearning and Software for Education (eLSE)* (Bucharest), 133–141.
- Frippiat, D., and Marquis, N. (2010). Web surveys in the social sciences: an overview. *Population* 65, 309–338. doi: 10.3917/popu.1002.0309
- Gelinas, L., Pierce, R., Winkler, S., Cohen, I. G., Lynch, H. F., and Bierer, B. E. (2017). Using social media as a research recruitment tool: ethical issues and recommendations. *Am. J. Bioethics* 17, 3–14. doi: 10.1080/15265161.2016.1276644
- Gilbert, S. B., Jasper, A., Sepich, N. C., Doty, T. A., Kelly, J. W., and Dorneich, M. C. (2021). “Individual differences task attention in cybersickness: A call for a standardized approach to data sharing,” in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (Lisbon: IEEE), 161–164.
- Grinyer, K., and Teather, R. J. (2022). “Effects of field of view on dynamic out-of-view target search in virtual reality,” in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (Christchurch: IEEE), 139–148.
- Hansen, J. P., Rajanna, V., MacKenzie, I. S., and Baekgaard, P. (2018). “A fitts’ law study of click and dwell interaction by gaze, head and mouse with a head-mounted display,” in *Proceedings of the Workshop on Communication by Gaze Interaction, COGAIN ’18* (New York, NY: Association for Computing Machinery).
- Herrera, F., Bailenson, J., Weisz, E., Ogle, E., and Zak, J. (2018). Building long-term empathy: a large-scale comparison of traditional and virtual reality perspective-taking. *PLoS ONE* 13:e0204494. doi: 10.1371/journal.pone.0204494
- Howarth, P. A., and Hodder, S. G. (2008). Characteristics of habituation to motion in a virtual environment. *Displays* 29, 117–123. doi: 10.1016/j.displa.2007.09.009
- Iacono, V. L., Symonds, P., and Brown, D. H. (2016). Skype as a tool for qualitative research interviews. *Sociol. Res. Online* 21, 103–117. doi: 10.5153/sro.3952
- ISO (2000). “ISO 9241-9, Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: requirements for non-keyboard input devices,” in *International Organization for Standardization*.
- Jacob, R. J. K., Girouard, A., Horn, M. S., and Zigelbaum, J. (2008). “Reality-based interaction: a framework for post-WIMP interfaces,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY: ACM). doi: 10.1145/1357054.1357089
- Kizilcec, R. F., and Saltarelli, A. J. (2019). “Cues impact women’s participation in STEM education,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–10.
- Koeman, L. (2018). How many participants do researchers recruit? a look at 678 ux/hci studies - lisa koeman - blog. Available online

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



at: <https://lisakoeman.nl/blog/how-many-participants-do-researchers-recruit-a-look-at-678-ux-hci-studies/> (accessed May 19, 2021).

Koeman, L. (2020). *Hci/ux research: what methods do we use?* - lisa koeman - blog. Available online at: <https://lisakoeman.nl/blog/hci-ux-research-what-methods-do-we-use/> (accessed May 19, 2021).

Kroma, A., Al Kassm, N., and Teather, R. J. (2021). "Motion sickness conditioning to reduce cybersickness," in *Symposium on Spatial User Interaction, SUI '21* (New York, NY: Association for Computing Machinery).

Kyian, S., and Teather, R. (2021). "Selection performance using a smartphone in vr with redirected input," in *Symposium on Spatial User Interaction, SUI '21* (New York, NY: Association for Computing Machinery).

LaViola J. J. Jr., Kruijff, E., McMahan, R. P., Bowman, D., and Poupyrev, I. P. (2017). *3D User Interfaces: Theory and Practice*. Boston, MA: Addison-Wesley Professional.

Lin, I. Y., and Kwantes, C. T. (2015). Potential job facilitation benefits of "water Cooler" conversations: the importance of social interactions in the workplace. *J. Psychol.* 149, 239–262. doi: 10.1080/00223980.2013.874322

Ma, X., Cackett, M., Park, L., Chien, E., and Naaman, M. (2018). "Web-based vr experiments powered by the crowd," in *Proceedings of the 2018 World Wide Web Conference, WWW '18* (Geneva: CHE. International World Wide Web Conferences Steering Committee), 33–43.

MacKenzie, I. S. (2013). *Human-Computer Interaction: An Empirical Research Perspective, 1st Edn.* San Francisco, CA: Morgan Kaufmann Publishers Inc.

MacKenzie, S. I. (1992). Fitts' law as a research and design tool in human-computer interaction. *Hum. Comput. Interact.* 7, 91–139. doi: 10.1207/s15327051hci0701\_3

Miller, M. R., Jun, H., Herrera, F., Villa, J. Y., Welch, G., and Bailenson, J. N. (2019). Social Interaction in augmented Reality. *PLoS ONE* 14, e0216290. doi: 10.1371/journal.pone.0216290

Mittelstaedt, J. M., Wacker, J., and Stelling, D. (2019). VR aftereffect and the relation of cybersickness and cognitive performance. *Virtual Real.* 23, 143–154. doi: 10.1007/s10055-018-0370-3

Moran, K. (2020). *Remote Usability-Testing Costs: Moderated vs. Unmoderated.* Nielsen Norman Group.

Moran, K., and Pernice, K. (2020). *Remote Moderated Usability Tests: Why to do Them.* Nielsen Norman Group.

Moss-Racusin, C. A., Pietri, E. S., van der Toorn, J., and Ashburn-Nardo, L. (2021). Boosting the sustainable representation of women in STEM with evidence-based policy initiatives. *Policy Insights Behav. Brain Sci.* 8, 50–58. doi: 10.1177/2372732220980092

NASA (2022). NASA TLX: Task Load Index. Available online at: <https://humansystems.arc.nasa.gov/groups/tlx/> (accessed May 10, 2020).

Nickerson, J., Corter, J., Esche, S., and Chassapis, C. (2007). A model for evaluating the effectiveness of remote engineering laboratories and simulations in education. *Comput. Educ.* 49:708–725. doi: 10.1016/j.compedu.2005.11.019

Ostafin, J. (2020). "do it yourself, user" - product design and research workshops. Available online at: <https://projectpeople.io/blog/do-it-yourself-user-product-design-and-research-workshops/> (accessed May 25, 2021).

Paternò, F., Russino, A., and Santoro, C. (2007). "Remote evaluation of mobile applications," in *Task Models and Diagrams for User Interface Design*, eds M. Winckler, H. Johnson, and P. Palanque (Berlin/Heidelberg: Springer Berlin Heidelberg), 155–169.

Pollock, J., and Teather, R. J. (2020). "Countmarks: multi-finger marking menus for mobile interaction with head-mounted displays," in *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings* (Berlin; Heidelberg: Springer-Verlag), 241–260.

Puma Applied Science (2017). Puma method for prevention of motion sickness. Available online at: <https://www.pumamethod.com/> (accessed May 23, 2021).

Puma, S. C., and Puma, S. W. (2007). *Method for conditioning to prevent motion sickness.* US Patent No. 7,288,057.

Ratcliffe, J., Soave, F., Bryan-Kinns, N., Tokarchuk, L., and Farkhatdinov, I. (2021a). "Extended reality (xr) remote research: A survey of drawbacks and opportunities," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21* (New York, NY: Association for Computing Machinery).

Ratcliffe, J., Soave, F., Hoover, M., Ortega, F. R., Bryan-Kinns, N., Tokarchuk, L., et al. (2021b). "Remote xr studies: Exploring three key challenges of remote xr experimentation," in *Extended Abstracts of the 2021 CHI Conference on Human*

*Factors in Computing Systems, CHI EA '21* (New York, NY: Association for Computing Machinery).

Samimi, E., and Teather, R. J. (2022a). "Impress: Improved multi-touch progressive refinement selection strategy," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (Christchurch: IEEE), 304–309.

Samimi, E., and Teather, R. J. (2022b). "Multi-touch smartphone-based progressive refinement vr selection," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (Christchurch: IEEE), 582–583.

Scavarelli, A., Arya, A., and Teather, R. J. (2019). "Circles: exploring multi-platform accessible, socially scalable VR in the classroom," in *2019 IEEE Games, Entertainment, Media Conference (GEM)* (New Haven, CT), 1–4.

Scavarelli, A., Arya, A., and Teather, R. J. (2020). Virtual reality and augmented reality in social learning spaces: a literature review. *Virtual Real.* 25, 257–277. doi: 10.1007/s10055-020-00444-8

Scheier, M. F., and Carver, C. S. (1985). The self-consciousness scale: a revised version for use with general populations. *J. Appl. Soc. Psychol.* 14, 687–699. doi: 10.1111/j.1559-1816.1985.tb02268.x

Schmidt, A., Alt, F., and Mäkelä, V. (2021). "Evaluation in human-computer interaction - beyond lab studies," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, CHI EA '21* (New York, NY: Association for Computing Machinery).

Shin, D. (2018). Empathy and embodied experience in virtual environment: to what extent can virtual reality stimulate empathy and embodied experience? *Comput. Human Behav.* 78, 64–73. doi: 10.1016/j.chb.2017.09.012

Spittle, B., Xu, W., Frutos-Pascual, M., Creed, C., and Williams, I. (2021). "Socially distanced: have user evaluation methods for immersive technologies changed during the COVID-19 pandemic?" in *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (Bari: IEEE), 415–420.

SREAL (2022). Planning a virtual experience research accelerator (VERA) 2022. Available online at: <https://sreal.ucf.edu/vera-summary/> (accessed April 20, 2022).

Steed, A., Frlston, S., Lopez, M. M., Drummond, J., Pan, Y., and Swapp, D. (2016). An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment. *IEEE Trans. Vis. Comput. Graph.* 22, 1406–1414. doi: 10.1109/TVCG.2016.2518135

Steed, A., Ortega, F. R., Williams, A. S., Kruijff, E., Stuerzlinger, W., Batmaz, A. U., et al. (2020). Evaluating immersive experiences during COVID-19 and beyond. *Interactions* 27, 62–67. doi: 10.1145/3406098

Teather, R. J., and Stuerzlinger, W. (2011). Pointing at 3D targets in a stereo head-tracked virtual environment. *3DUI 2011 - IEEE Symposium on 3D User Interfaces 2011, Proceedings*, 87–94. doi: 10.1109/3DUI.2011.5759222

The Ministry of Labour, Training and Skills Development (2022). Skilled Trades. Available online at: <https://www.ontario.ca/page/skilled-trades> (accessed May 10, 2022).

Townsend, E., Nielsen, E., Allister, R., and Cassidy, S. A. (2020). Key ethical questions for research during the COVID-19 pandemic. *Lancet Psychiatry* 7, 381–383. doi: 10.1016/S2215-0366(20)30150-4

Unity Technologies (2021). Unity pulse. Unity. Available online at: <https://unity.com/unity-pulse>

Usoh, M., Catena, E., Arman, S., and Slater, M. (2000). Using presence questionnaires in reality using presence questionnaires in reality using presence questionnaires in reality. *Presence* 9, 497–503. doi: 10.1162/105474600566989

W3C (2019). WebXR Device API. Available online at: <https://www.w3.org/TR/webxr/> (accessed January 1, 2020).

Wala, K., and Maciejewski, M. (2020). On the border of data, creation and design decisions. introduction to research and design workshops. Krakow: Faculty of Industrial Design of the Academy of Fine Arts. Available online at: <https://bit.ly/2RCTLHs> (accessed May 25, 2021).

Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Front. Psychol.* 10, 158. doi: 10.3389/fpsyg.2019.00158

WHO (2022). Coronavirus disease (COVID-19) pandemic. Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed April 11, 2022).

Wiberg, M., Taylor, A., and Rosner, D. (2020). Responding to the COVID-19 pandemic: an invitation. *Interactions* 27, 5. doi: 10.1145/3392538

XRDRN (2020). About XRDRN. Available online at: <https://www.xrdrn.org/about/> (accessed April 20, 2022).