Evaluation Techniques Used to Evaluate Extended Reality (XR) Head Mounted Displays (HMDs) Used in Healthcare: A Literature Review

Authors:

Dr. Payal Ghatnekar - Immersive Technology Research Development Lead, Digital Futures Lab, Torbay and South Devon NHS Foundation Trust (2023)

Dr. Mona Seyed Esfahani- Senior Lecturer in Communication, SIG Leader of Health and Innovation group in ISPIM, Bournemouth University (2023)

Digital Futures lab is an immersive technology lab based at the Torbay and South Devon NHS Foundation Trust (Torbay Hospital). Our lab is one of the first immersive technologies research spaces, established in 2016. Digital Futures work is supported by Health Education England (HEE) to inform national strategies for development, implementation, upscaling and research in immersive technologies. At Digital Futures lab we use different immersive technologies to create interventions that meet the requirements of Torbay hospital's key priorities- healthcare workers, medical students and patients. Our aim is to inspire and encourage discussions on the use and evaluation of immersive technologies in healthcare. To learn more about work, please visit **immersive.tsdft.uk**

NHS Torbay and South Devon NHS Foundation Trust

DIGITAL FUTURES HUMAN CENTERED DIGITAL INNOVATION

DIGITAL FUTURES HUMAN CENTERED DIGITAL INNOVATION

Summary

Extended Reality (XR) Head Mounted Displays (HMDs) are used across various healthcare pathways for staff/student education and training, and for improving patient experiences. As XR HMDs become affordable, accessible and their acceptance increases, it is critical to document the techniques used for evaluating the technology, processes of user engagement and immersion, and outcomes. At present there is limited research on evaluation techniques used to evaluate XR HMDs.

This manuscript presents findings from 104 clinical studies that use XR HMDs. The aim of this review is to give the user an insight into the current healthcare XR HMD landscape by presenting the different HMDs used, variety of XR interventions and their applications across medical pathways and intended research outcomes of the XR applications. The manuscript further guides the reader toward a detailed documentation of evaluation techniques used to investigate antecedents and consequences of using XR and delivers a critical discussion and suggestions for improvement of XR evaluation practices. This paper will be of excellent use to clinicians, academics, funding bodies and hospital decision makers who would like suggestions for evaluating the efficacy and effectiveness of XR HMDs. The authors hope to encourage discussions on the importance of improving XR evaluation practices.

Contents

Introduction

Healthcare education, training and patient experiences have undergone revolutionary transformation over the past two decades as a result of wide spread adoption of immersive virtual technologies in healthcare education and training are growing rapidly. However, the research investigating evaluation methods within the context of healthcare XR HMDs is still at an early stage.

XR technology is still new, and access is limited to healthcare and educational organisations that are pioneers in this space. There may be healthcare educators and practitioners who are keen on adopting XR and may require additional information to help them make their decision. Moreover, many educators and practitioners may be questioning the scalability of XR across other pathways and need guidance on evaluation metrics to measure research outcomes. To establish an understanding of the current XR HMD based interventions and evaluations landscape, this research aims to investigate existing literature and answer the following research questions,

1. What kind of XR devices have been used in literature?
2. What are the XR interventions and which healthcare pathways have they been used?
3. What are the intended research outcomes of XR applications?
4. What are the evaluation techniques used to investigate the antecedents and consequences of using XR in healthcare?

To answer the research questions, this paper presents findings from a short literature review of healthcare XR HMD based interventions used for education and training of healthcare workers/students, and improvement of patient experiences. This literature review is based on a search conducted on the PubMed database. Search was conducted for articles published between 2012-2020. Accessible standalone HMDs like Google Cardboard and Samsung Galaxy VR started to gain popularity within healthcare education and training during early 2012, making this a reasonable year to set search range. XR HMDs have existed prior to 2012, with significant use in industries like aeronautics, defense, and computing/gaming. However, there is a noticeable trend that shows a rise in healthcare publications involving XR HMDs post 2011. Search terms used were broad, to encompass all possible interventions that may fall under the definition of XR HMDs- *Virtual Reality, VR, Augmented Reality, AR, Mixed Reality, MR, Extended Reality, XR, Head Mounted Displays, HMD, including names of popular headsets HoloLens, Oculus, HTC Vive*. These terms were searched in population settings for *healthcare, medical, medicine, surgical* and *surgery*.

Articles written in English have been included in the review. Interventions used in this review range across patients (adults and children), and healthcare workers and students' education and/or training. This paper does not include any grey literature. A total of 104 articles were selected as a final sample.

The characteristics of the articles selected for this review were recorded using a template with the categories- *Title, Authors, Journal, Publication Year, Intervention, Type of Intervention,* *Sample and Research Design, XR Device, Measurements/Metrics, Result, Conclusion* and *Additional Notes.* Coding in this manner has helped in exploring all aspects of the interventions, providing a better understanding of the XR healthcare landscape.

Principle Findings

XR HMDs in the market

At present there are a variety of XR headsets in the global market, ranging from high-end expensive headsets like the Microsoft Hololens, Varjo, HTC Vive Focus to the budget friendly and more accessible Meta (previously Oculus) Quest and Pico Neo. Headsets like the Quest and HTC Vive use VR technology that transports users into entirely virtual worlds, where their physical reality is blocked out. Microsoft HoloLens and Moverio BT-200, use AR technology with see-through glasses that display augmented layers placed on physical reality. Headsets like the Varjo XR 3 combine AR and VR technologies to provide a blended experience with pass through between physical and digital world. AR and VR technologies have been on their unique development paths. However, lately there appears to be a greater interest in convergence between both worlds with the Quest and Pico VR headsets also focusing on delivering AR experiences like the Varjo.

Higher end headsets that provide significantly greater quality of experiences, like the Varjo and Valve Index are tethered and require a VR ready computer to run the applications. Others like HTC Vive Focus 3, Meta Quest and Microsoft HoloLens are untethered. Older headsets like the Google Cardboard and Samsung Gear VR (pre-2021), require smart phones to deliver immersive VR experiences. The XR HMD market is under continuous transformation and able to deliver more usable, affordable, and technologically advanced HMDs. Meta Quest and Pico Neo are good examples of headsets becoming affordable, accessible, ergonomic, and capable of delivering immersive experiences of a higher quality. Most headsets mentioned require hand-held controllers for movement through virtual worlds, which have a dual function of providing haptics feedback. However, the HoloLens relies on user's hand gestures and the Varjo incorporates both hand gestures and eye movements. With the recent updates on the Quest 2, there appears to be a growing interest in hand and eye-tracking for user movement. Furthermore, technologies such as full body trackers and omni-directional treadmills can be paired with the available XR HMDs to enhance the immersion in the virtual worlds.

XR HMDs and pathways in healthcare literature

Within selected healthcare literature, 29 different XR HMDs that provide a mix of VR and AR experiences have been used (Figure 1). Seven papers do not mention the name of the XR HMD used but they have been included in the research for insights into evaluation techniques. Approximately 26% of the selected studies use Microsoft Hololens, followed by Oculus Rift (23%) and, Samsung Gear VR and HTC Vive (9% each). Microsoft Hololens is an AR HMD, Oculus Rift 1 and 2, Gear VR and HTC Vive are VR HMDs. For additional information on headsets corresponding with the authors, please refer to Table 1 in Supplementary Material.

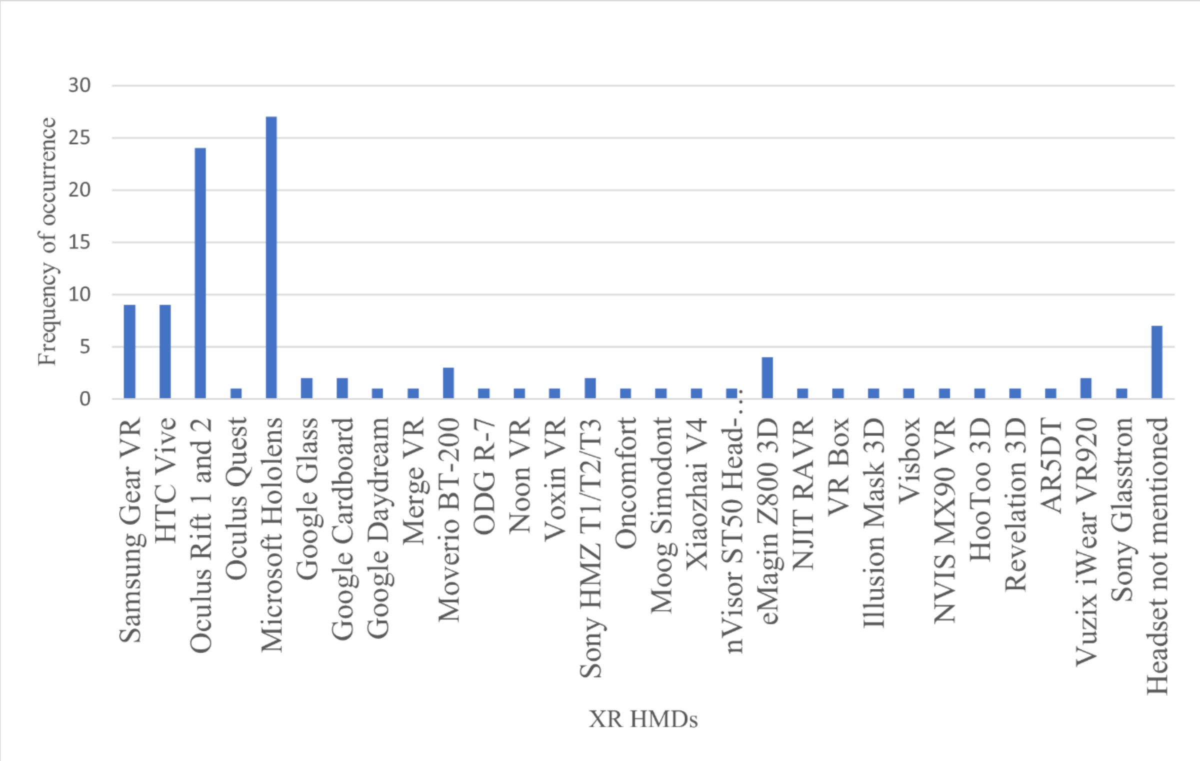

Figure 1: HMDs used in literature.

Figure 2 displays healthcare pathways that have used the more commonly used headsets. The size of text displays the frequency of use of the headset in that pathway. Microsoft Hololens has been used frequently in the pathways of Neurology, followed by Cardiology and Urology. Oculus Rift 1 and 2 have been used commonly in Psychology and Neurology. Samsung Galaxy VR is often used in Psychology. HTC Vive is used in Psychology and Orthopaedic pathways.

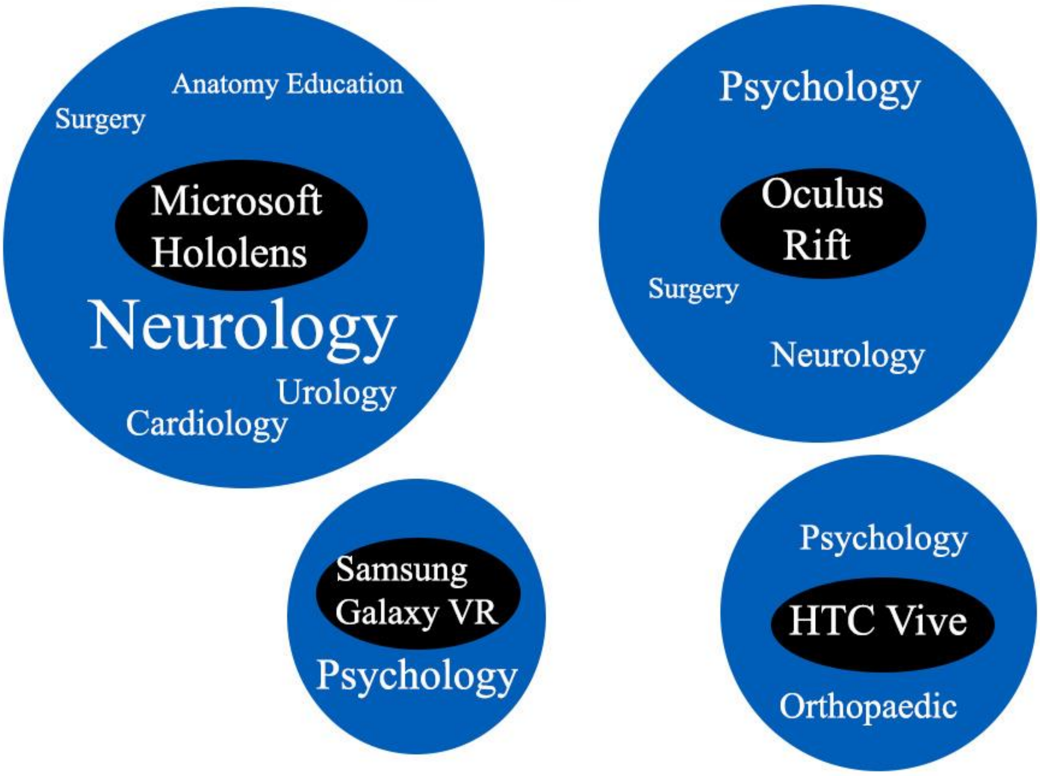

Figure 2: Popular headsets and pathways with frequent use.

Figure 3 shows the pathways and intervention types found in the selected literature. Detailed information about the interventions has been compiled under 26 different healthcare specialisations in Table 2 (Supplementary Material). There are two categories under which the interventions fall- healthcare staff and students, and patients. The healthcare staff and student interventions are aimed toward education and training. Patient interventions are for educating patients and distracting them from pain and anxiety. Majority of the interventions in the selected literature focus on patient interventions (62%). The remaining 36% are interventions in the domain of healthcare workers and students. Two interventions fall in the category of both, patients and healthcare workers and students' domains.

As Figure 3 shows, the greatest number of healthcare staff and student education and training interventions can be found in the Surgery pathway (7%). Large variety of patient-based interventions can be found in Neurology (12%), Psychology/Psychiatry (19%) and Nursing (6%).

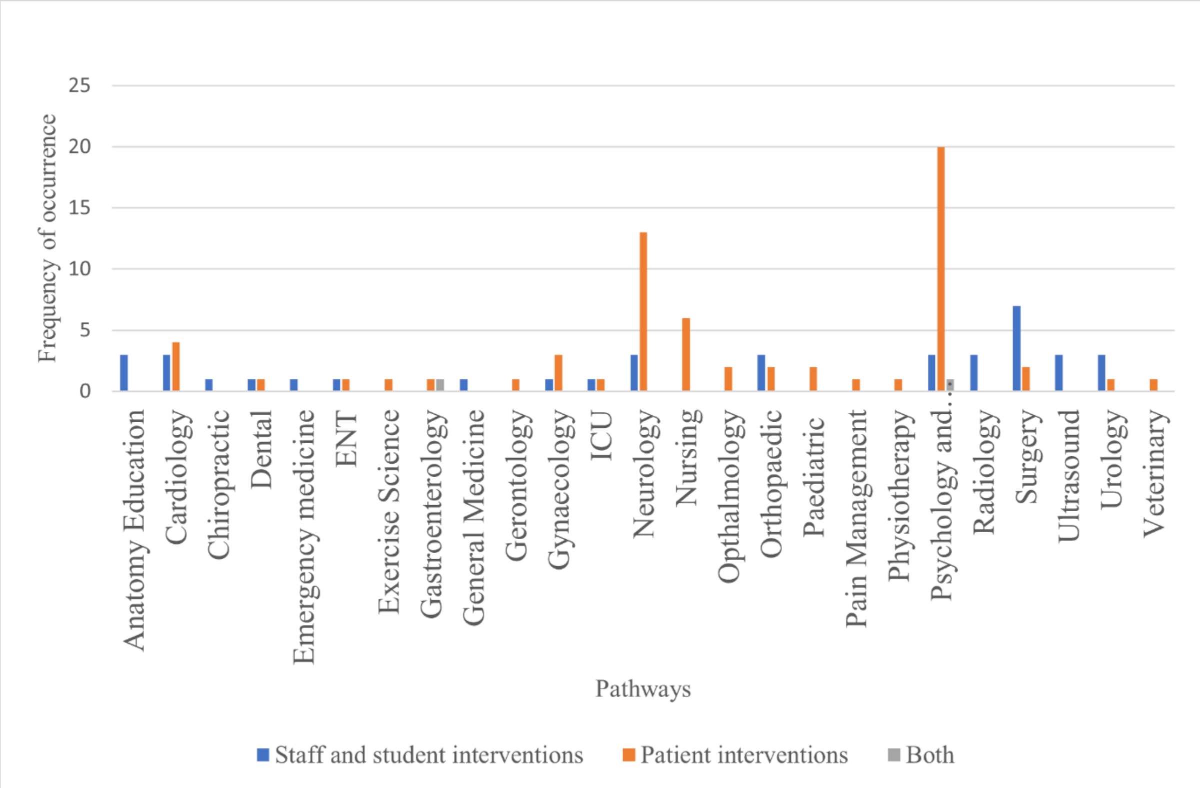

Figure 3: Intervention pathways, domains, and frequency of occurrence.

Evaluation metrics

Figures 4 and 5 show that priorities for both domains differ, and so do evaluation techniques. Please refer to Table 3 in Supplementary Material for additional information on domains, pathways, authors, outcomes, and evaluation techniques used. In the domain of healthcare workers and students (Figure 4), XR HMDs are often used to educate and evaluate technical competencies and knowledge retention. Other areas of evaluation are usability, user experience (UX), feasibility, acceptability, applicability, motivation, and confidence gained after using XR HMDs.

Figure 5 shows that the key evaluations in the patient's domain focus on pain, anxiety and fear distraction, improvement in post procedure outcomes, usability, UX, cybersickness, user immersion and presence in the virtual experience, and cost-effectiveness.

In this review, evaluation types within selected literature have been categorised under three themes based on antecedents and consequences of XR, 1. HMD technology and XR experience, 2. the process of immersion and engagement afforded by the HMD and XR experience, and 3. the evaluation of the actual outcomes. The categorisation results will assist in gaining better understanding of the evaluation landscape, especially the gaps in healthcare XR HMD evaluation literature. Figures 4 and 5 present the types of evaluations (rectangle shape) and techniques used to evaluate (ellipse and circle shapes). The size of the text in the ellipse represents the frequency of use of the evaluation technique.

Evaluations in Healthcare workers and students domain

Figure 4 presents evaluation techniques used in healthcare workers and students domain. Approximately 71% of the studies in the healthcare workers and students domain have focused on evaluations that can be categorised under technology evaluations. The System Usability Scale has been used in two studies, along with custom designed opinion surveys to investigate user experience and usability. Suitability, risk, cybersickness, accuracy, acceptability and feasibility are evaluated through custom questionnaires (Table 3 in Supplementary Material). To evaluate motivation, the validated scale Instructional Materials Motivation Survey (IMMS) has been used once. Largely, qualitative approaches have been used to conduct technology user assessments.

Process evaluation in this domain focuses on investigating presence, confidence, and motivation as antecedents to outcomes (Figure 4). Variety of custom opinion surveys are used to investigate various aspects of processes that may contribute toward users engagement in the experience. However, within the healthcare workers and students domain, only 10.5% of the total studies focus on factors that can be categorised under process evaluations.

Approximately 71% of the evaluations also fall in the category of outcomes evaluation. The validated NASA Task Load (NASA -TLX) questionnaire has been used multiple times (8%) to assess the task, system, and user’s performance. Another 8% of the studies use Objective Structure Assessment of Technical Skills (OSATS)/Objective Structured Clinical Examination (OSCE)/Global Rating Scales (GRS) to assess technical competencies. These scales are popular within healthcare education and used commonly. Custom questionnaires and assessments have been used to evaluate pathway specific technical competencies and knowledge retention. Mostly quantitative approaches have been used to evaluate outcomes.

Evidence from the literature review shows that the focus of research in the domain of healthcare staff and students is largely on technology and outcomes evaluations. Technology evaluations mostly use a qualitative custom questionnaire approach. Within healthcare, outcomes evaluations are preferred as they provide quantitative task specific metrics. However, the low focus on process evaluations can be noticed (Please refer to Table 3 for additional information). Factors that act as antecedents to outcomes are not investigated clearly. The domain will also benefit from using mixed methods instead of focusing solely on either qualitative custom questionnaires or quantitative metric specific data.

Evaluations in patients domain

Figure 5 presents the different evaluation techniques used to evaluate XR interventions in the patients domain. In the category of technology evaluations, approximately 45.5% of the evaluate factors such as user experience and usability, acceptance, cost-effectiveness of implementing XR, and cybersickness experienced by patients while using the XR HMDs. Patients’ UX and usability of the XR HMD and experience has been measured using validated scales, Quality of Recovery Survey (QoR-40) and Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). Several other studies have developed custom questionnaires to study patients' opinions. Questions based on the Technology Acceptance Model (TAM) have been used by one study to investigate user acceptance, but the popular method of assessing acceptance is through custom questionnaires. For evaluating user motivation, the Intrinsic Motivation Inventory has been used by two studies. Approximately, 16% of the studies evaluate user cybersickness, 13% of these use the Simulator Sickness Questionnaire (SSQ). Others use custom simulator sickness questionnaires to measure the same (Table 3 in Supplementary Material).

Process evaluations in patient literature focus on immersion, and presence. Approximately 13% of the studies have investigated user presence, and two of these studies also investigate immersion. For assessment of user's presence and immersion, the validated Witmer and Singer Presence Questionnaire, Gatineau Presence Questionnaire (GPQ), and the International Television Commission Sense of Presence and Inventory scales have been used.

About 41% of the studies in the patient domain literature use XR to focus on outcomes of pain and anxiety reduction. As this section contains many scales, Figure 5 only includes those scales which have been referenced multiple times. The Visual Analogue Scale (VAS) is a validated scale that has been used often within literature, followed by the Graphic Rating Scale (GRS), Numeric Rating Scale, State Trait Anxiety Inventory (STAI), Amsterdam Preoperative Anxiety and Information Scale, Anxiety Sensitivity Index 3 (ASI-3), and Inventory of Depression and Anxiety Symptoms (IDAS). Researchers have also developed custom questionnaires to evaluate changes in pain and anxiety levels pre-and-post XR intervention. For paediatric patients' pain and anxiety assessment, validated scales such as the Childhood Anxiety Sensitivity Index, Coloured Analogue Scale (CAS), Children's Fear Scale, Child-rated Faces Pain Scale-Revised (FPS-R), Modified Yale Preoperative Anxiety Scale -Korean version of mYPAS and Wong and Baker FACES scale are used. These scales are used commonly in healthcare as they are easy for younger patients to understand.

Approximately 50% of the studies evaluate patient outcomes. Most studies have used customised outcome testing questionnaires to evaluate patient's education and skills of symptom management. One study has used the validated NASA Task Load Index (NASA-TLX) to assess task training. This scale is also used in healthcare workers and student's domain for skills assessment.

Within patients domain, there is a wider variety of evaluation techniques. One of the primary reasons for this is the greater use of XR for patient interventions. The various validated scales and custom questionnaires used for technology, process, and outcome evaluations, generate a mix of quantitative and qualitative data. Immersion and presence are evaluated as antecedents to patient behaviour change (mostly pain and anxiety distraction). Application of validated scales for the evaluation of immersion and presence show that these complex topics can be investigated. As evidenced from the selected literature, there is an opportunity to improve process evaluations in both patient and healthcare workers and students domains.

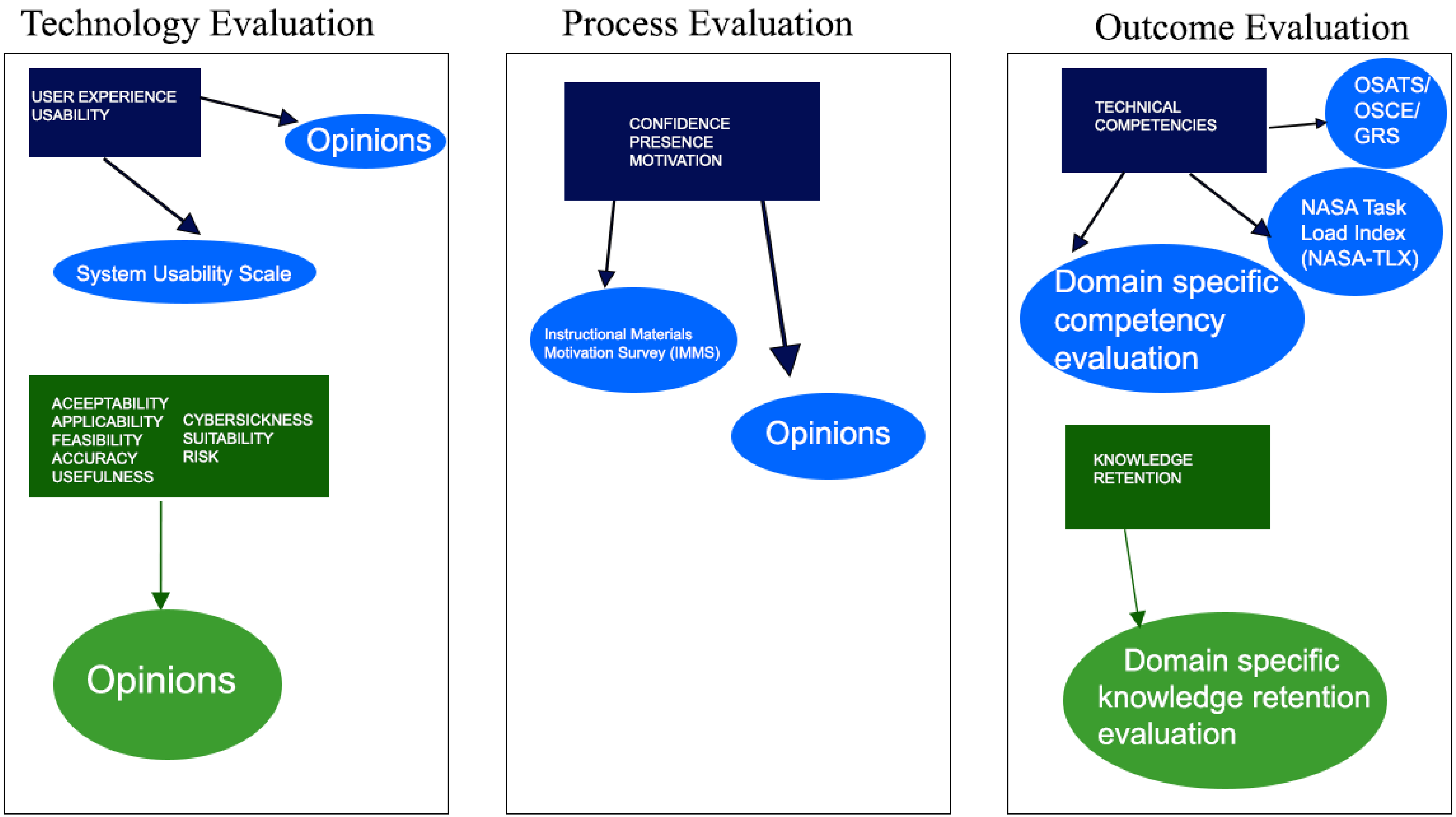

Figure 4: Healthcare workers and students XR HMD evaluation

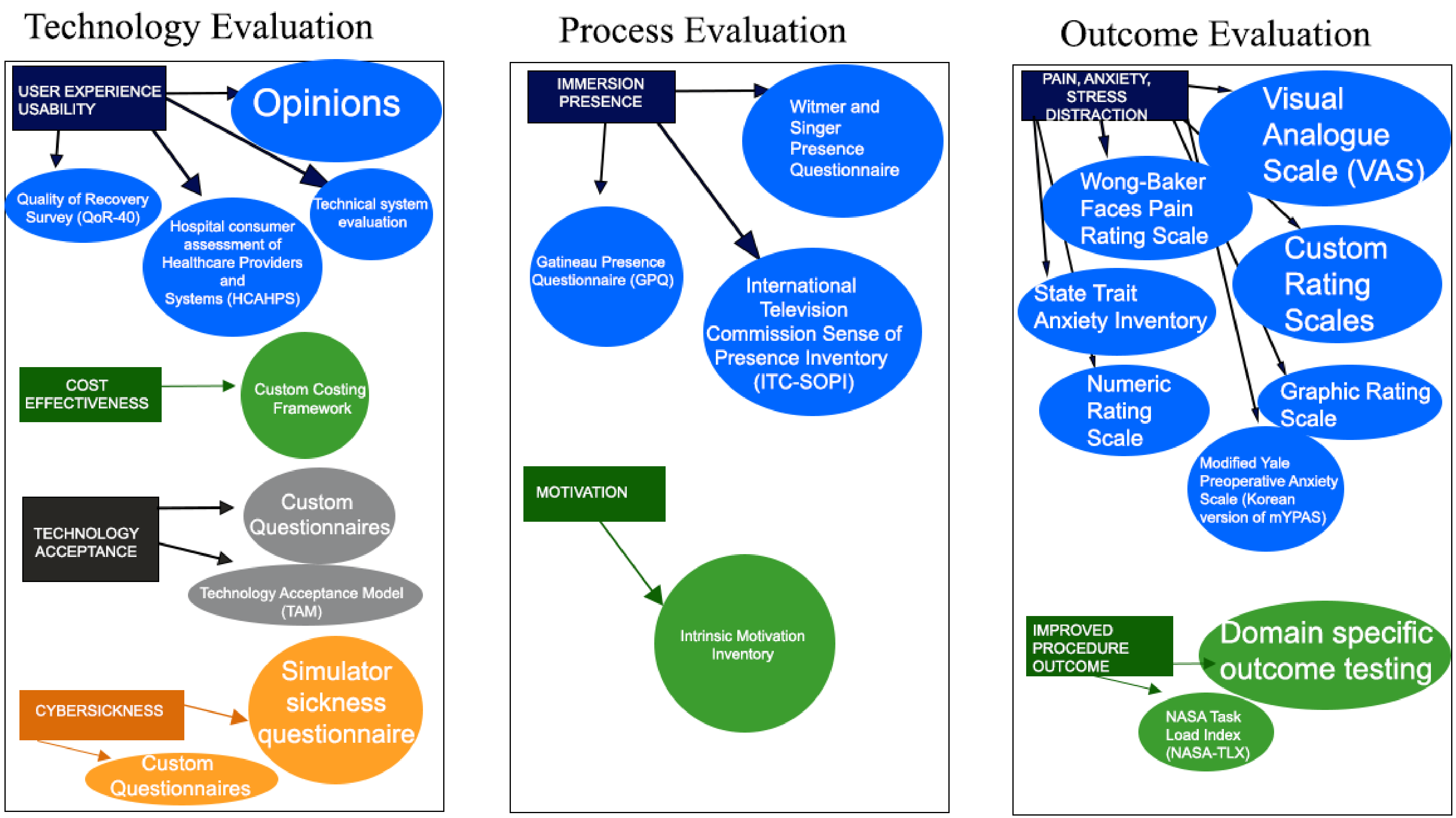

Figure 5: Patient XR HMD evaluations.

Summation of evaluation practices in both domains

Evaluation techniques presented in this literature review (Table 3) suggest a variety of quantitative, qualitative and mixed methods approaches. Often, researchers' pair multiple quantitative and qualitative tools to draw a better understanding of the effectiveness, efficacy, feasibility, usability, and user experience of XR experiences and HMDs. A variety of different aspects of the XR HMD technology, process of engagement and immersion and intervention outcomes are evaluated in the literature. These categories are introduced by the authors with the aim to assist the reader in understanding the gaps in evaluation literature.

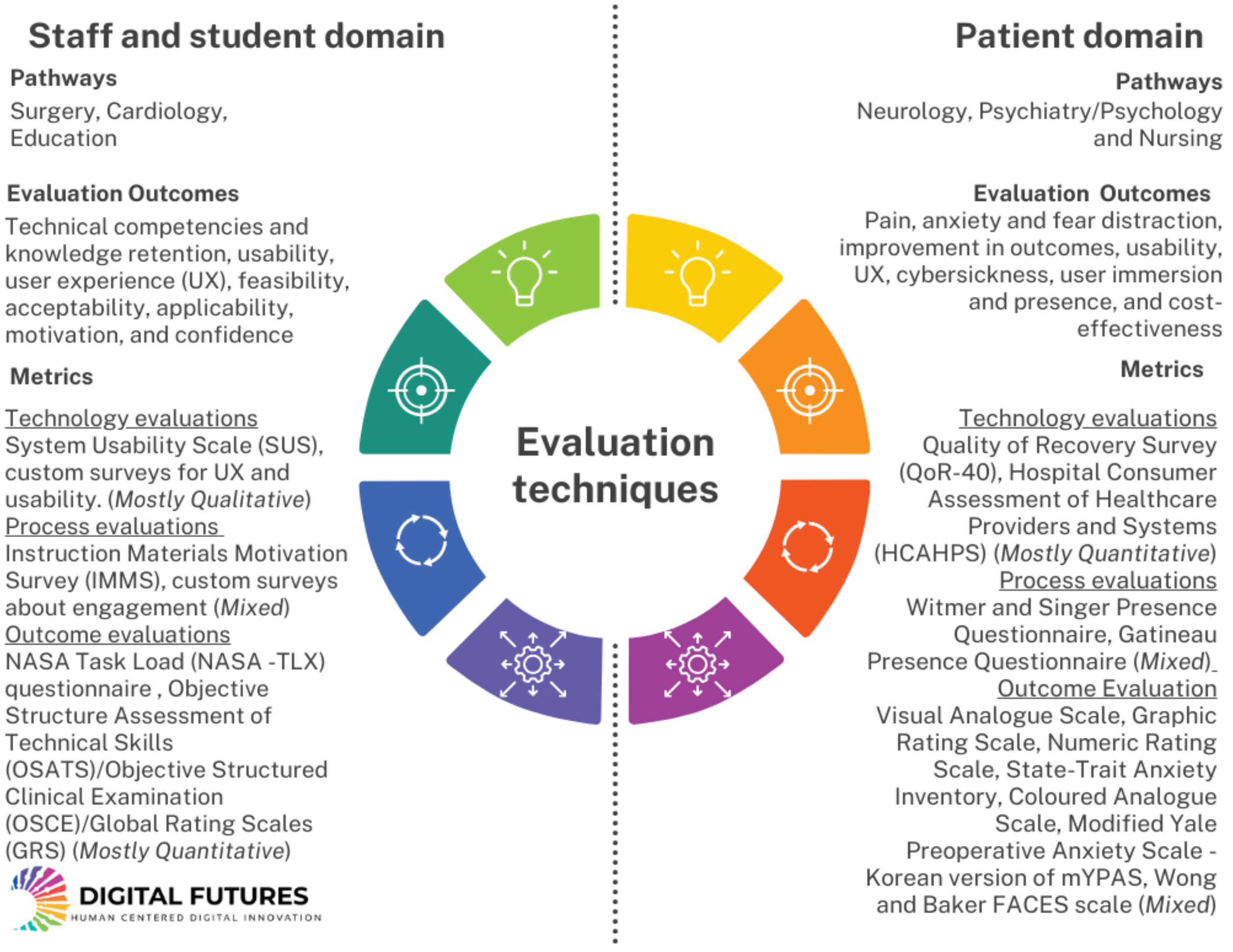

Figure 6: Summation of evaluation practices

The one theme that repeats itself across studies is that of the emphasis on outcomes of the interventions, in both domains. This is valid; however, such evaluations present a very narrow perspective. Equally, there appears to be an interest in investigating the technology context through users' perceptions and opinions about the XR HMD user experience and usability. These investigations are pursued in both domains. However, enquiries into the role of process through evaluations that investigate user's presence, engagement, immersion, or any other antecedents to the intervention, are less commonly pursued. Specifically, in the domain of healthcare workers and students, process evaluations are rarely pursued. When process is investigated, generally, a custom designed questionnaire has been used instead of validated scales. Some presence and immersion scales are used in the domain of patients. However both domains may benefit from higher quality investigations into understanding how users engage with the XR HMD, factors that affect the engagement and/or experience to facilitate intervention outcomes. Presence and immersion may play an important role. These two factors may be linked to better task performance; however, most evaluations do not present an in-depth picture of this relationship.

Discussion

XR HMDs are used quite widely by healthcare educators and practitioners. In the past decade, XR HMDs have undergone immense transformation, which has made the headsets affordable, usable, and highly experiential. Several literature reviews and systematic reviews have been published, all documenting various uses of VR technology across different healthcare pathways [13,40,41]. However, there still exists a gap in healthcare XR HMD evaluation literature. Furthermore, XR HMDs may still only be accessible to few healthcare workers and educators who are pioneers in technology adoption, and more importantly may have access to larger funds. A stronger evaluation evidence base is required to build use cases that can inspire, provide measurable data on the effectiveness and efficacy, and present additional information to guide investment decisions.

This review aims to provide answers to questions that may be on the minds of healthcare educators and practitioners who have not used XR or they may be in the early stages of using XR. This paper provides a glimpse into healthcare XR literature by condensing information regarding the various XR HMDs used, the interventions and pathways where they have been applied, the research outcomes pursued and most importantly, the evaluation tools and metrics used.

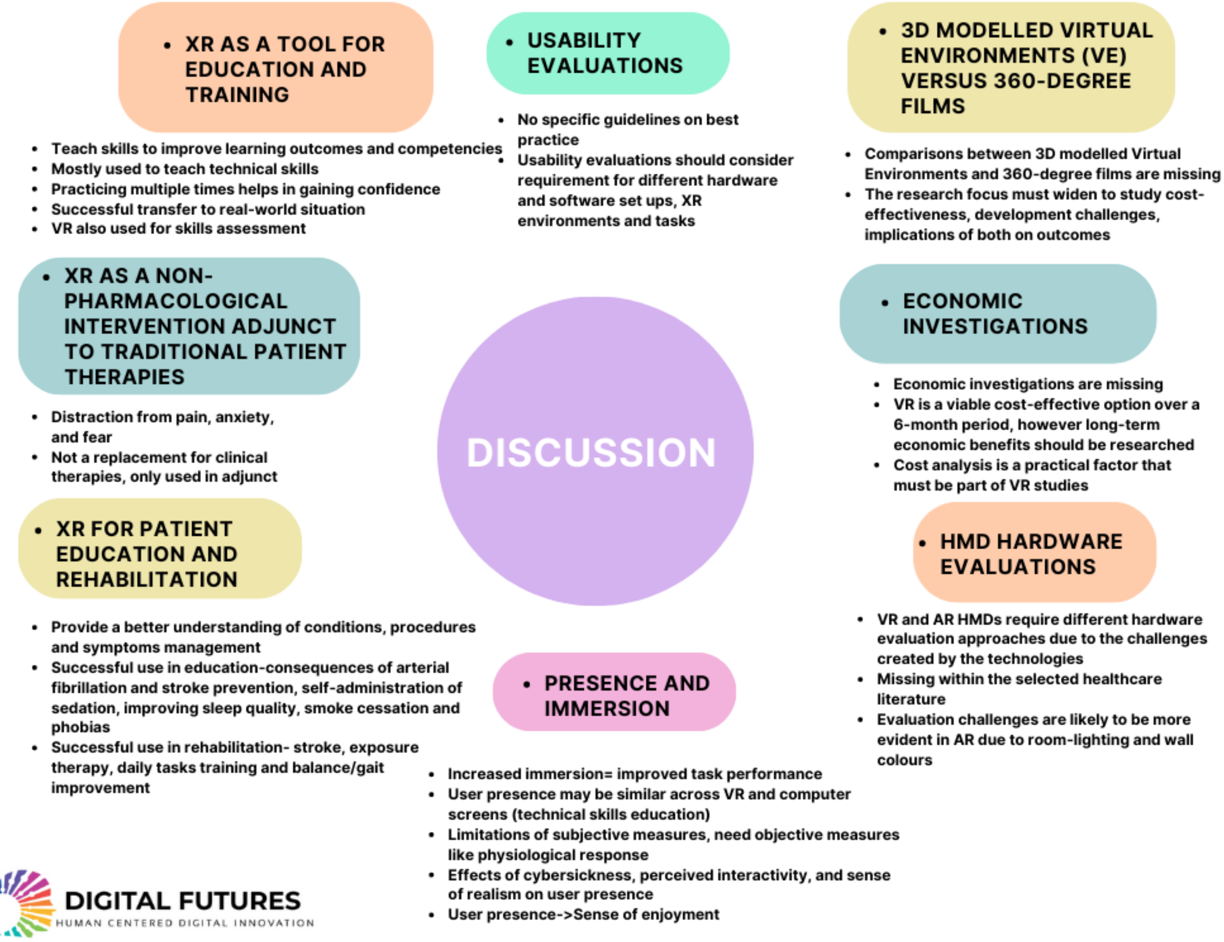

Figure 7: Discussion Points

XR as a tool for education and training of healthcare workers and students

The findings from this paper show that XR HMDs are used in healthcare workers and students domain, to teach technical skills, improve learning outcomes and competencies. Studies selected for this paper mostly focus on technical skills, but XR can also be used to teach non-technical skills [4] of communication, empathy and decision making. HMDs deliver immersive experiences that let users operate in virtually realistic, simulated environments with no real-world consequences [5]. Users can practice their skills multiple times until they gain confidence and can successfully transfer the knowledge to real-world situations [42,43]. Along with teaching technical skills, findings (Table 2) show that XR HMDs can also be used for assessing users skills[17,44]. Figure 3 shows that Surgery and Urology are two pathways with significant and varied use cases for XR HMDs for skills training and assessment.

XR as a non-pharmacological intervention adjunct to traditional patient therapies

XR HMDs are used to improve patient's experiences before, during and after clinical procedures. Findings from this research show that XR HMDs are popular for treating patients (including children as patients) pain and anxiety, pre-post procedures. Experiences offered through XR HMDs have been used successfully to distract patients from pain, anxiety, and fear[20,32,45]. XR HMDs are not a replacement for clinical therapies, but they may work in adjunct to these therapies, contributing toward improvement of patient outcomes.

XR for patient education and rehabilitation

Immersive visuals offered through XR HMDs can engage users and help them gain a better understanding [43,46] of their condition, procedures, management of symptoms etc. Findings from this research (Table 2) show that XR interventions are used to educate patients about their health conditions (e.g. consequences of arterial fibrillation and stroke prevention, self-administration of sedation, improving sleep quality, smoke cessation and phobias). XR HMDs have been used successfully for physical rehabilitation and psychological therapies [47,48]. Rehabilitation is another area where XR HMDs have been used in literature, specifically in stroke rehabilitation, exposure therapy, daily tasks training and balance/gait improvement. There are countless applications of XR in patient domain, however for them to be scaled up and adopted nationally, long term empirical evidence is required[16].

Presence and Immersion

Drawing upon Witmer and Singer's presence questionnaire [17] it has been suggested by [18] that increased immersion improves task performance. According to [39,50] immersion can be considered a quantified description of virtual reality. Immersion considers the vividness, extent, and inclusiveness of the technology. On the other hand, presence is the user's state of consciousness, it is the psychological sense of existing in virtual reality[18]. Users who are highly present are likely to become more engaged in the experience. The highest degree of presence is attained when the user behaves in the exact same manner as they would in everyday reality. Presence itself may lead to higher immersion, however [18]believe that presence may not directly be associated with improved task performance in Virtual Environment (VE).

Within the selected literature,[20] compared involvement of presence in motor skills transfer between a computer screen and VR HMD. The findings suggest that presence was similar across both devices. In the paper, [20] explain that user performance in VR environment may be related to levels of presence, and it could suggest a better transfer of skill acquisition. The limitations of subjective measures of presence have been documented by [20], highlighting the need to use objective measures such as physiological responses. A similar view has been stressed by [21]. Paired with a larger sample size, physiological measures may present a detailed analysis of presence during VR exposure. A potential link between presence and experience enjoyment has been documented by [22]. Along with an investigation of the implications cybersickness may have on user presence, [23] have also studied presence and sense of enjoyment. Findings from [23] suggest that perceived interactivity and realism may help in predicting spatial presence, there by showing them to be strong predictors of enjoyment. It has been determined by [24] that the sense of spatial presence is affected by interactivity and realism, and directly influences enjoyment of a VR experience.

Quality of the user interface may also independently predict the improvement in motor skill transfer, supporting the involvement of presence in transfer of skills [20]. When a user is fully present in the XR environment, there are fewer things that may distract them, thereby forcing the user to concentrate on what they are visualising, hearing and acting on with the help of haptics [25]. Within this literature review, Witmer and Singer's presence questionnaire [17], Gatineau presence questionnaire [21] and the International Television Commission Sense of Presence Inventory [26] are used to investigate the role of XR on user presence. It has been suggested by [27] that measures of presence from Witmer and Singer's questionnaire might contribute toward evaluation in terms of realism and overall UX, but it cannot help in diagnosing design flaws for formative evaluation. The Slater, Usoh and Steed presence questionnaire [18] is a set of heuristics and evaluation method, similar to Nielsen's heuristics, which can help in evaluating UX design as part of presence measurement.

Increased levels of immersion and presence suggest that a person is less likely to be able to distinguish between the real and the virtual world. Immersion and presence in the context of XR can be described as how closely the experience reflects the real world. Furthermore, higher sense of presence, lowers the risk of experiencing cybersickness [25] It has been suggested by [28] that immersion must increase user motivation, however, other aspects such as usability, usefulness, and acceptability should also be evaluated, alongside.

Findings from this review show limited investigations into processes that nudge and support users in achieving or completing intended outcomes. To improve XR evaluation practice, it is essential that future studies expand their research designs beyond investigations of the technology and outcomes by applying mixed methods approaches to study antecedents such as immersion and presence.

Usability evaluations

Findings from this literature review present several usability evaluations, which are conducted using custom questionnaires or opinions of users. In terms of validated scales, the System Usability Scale (SUS) [29] has been used by [9,31]. Literature on XR usability evaluations focuses on heuristics based approaches. Nielsen's heuristics may be a good fit to address usability concerns during early development stage of VR [32]. Nielsen's heuristics are ten heuristic evaluations (visibility of system status, match between system and real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, help users recognise, diagnose and recover from errors, help and documentation) from usability engineering. Small group of usability experts evaluate the user interface using these ten guidelines. Nielsen's ten heuristics have been extended by [27] who have added two more VE specific principles designed by [33]VE specific heuristics recommended by [27]are natural engagement, compatibility with the user's task and domain, natural expression of action, close coordination of action and representation, realistic feedback, faithful viewpoints, navigation and orientation support, clear entry and exit points, consistent departures, support for learning, clear turn taking, sense of presence. Usability testing of different XR environments is crucial to the success of the interventions. Designers competent in HCI knowledge could address usability issues early on in during the development phase [27], [33].

At present no specific guidelines exist on the best practices for usability testing of XR HMDs. A variety of different XR HMDs are currently in use. Future usability testing must take into consideration the different hardware and software set ups, along with additional usability research into XR environments and tasks. This requires an inclusive co-creation driven approach, which will take into consideration views of different users of the technology.

HMD hardware evaluations

Usability evaluations also tie into technical hardware and software evaluations, which are missing within the selected healthcare literature. Visual performance of the HMD display is key to an immersive experience [34]. However, evaluation of image quality is a challenge because of the diversity of XR HMDs, rapidly evolving technology and frequent software updates. Parameters such as luminance, contrast, field of view, refresh rate, temporal and spatial resolution, dynamic range, and connectivity can all affect the quality of the image [34,35]. Deciding upon a suitable image quality evaluation methodology would depend on the specific hardware technology [35]. Evaluation challenges are likely to be more evident in AR due to room-lighting and wall colours. Ambient lighting can impact the contrast and colour perception [35], making AR best suited as a tool that can be used in adjunct with traditional methods. Both, VR and AR HMDs require different hardware evaluation approaches due to the challenges created by the technologies. Focusing on hardware evaluations can help in identifying and improving HMD hardware through the creation of standards that HMD makers need to incorporate in the design process [34] Furthermore, healthcare researchers could explore computer sciences literature to borrow knowledge on best practices for hardware evaluations.

Economic investigations

Further analysis of the evaluation metrics used makes it evident that economic investigations on the cost-effectiveness, cost-benefit, and cost-utility of XR HMDs is missing in literature. For public health organisations like the National Health Service (NHS) UK, it is important to know the benefit of investing in XR technology. This data will be useful for mid to lower income countries as well, where educators may need to present a stronger data driven case to justify the use of XR HMDs. Within literature, [36] have suggested the importance of establishing economic value of XR, which is critical to its wide spread adoption. A cost-effectiveness analysis of VR for CBT in patients experiencing paranoid delusions has been conducted by [37]. The authors found that VR is a viable cost-effective option over a 6-month period, however long-term economic benefits should be researched. Similarly, [21]have studied cost-effectiveness of VR exposure in CBT for Social Anxiety Disorder (SAD). The authors have mentioned that cost analysis, is a practical factor that must be part of VR studies.

Economic analysis studies are lacking within the papers selected for this review. A theoretical framework has been developed by [38] for hospital administrators to evaluate Return on Investment (ROI) associated with implementation of VR. Specific to opioid utilisation, the authors analysed cost-savings from reduction in demand to reduction in length of stay, along with increased patient satisfaction. The authors found potential cost-savings may occur when length of stay is reduced, however, in isolated events of reduction in opioid utilisation, cost-savings may not make up for costs associated with implementing VR.

3D modelled Virtual Environments (VE) versus 360-degree films

Comparisons between 3D modelled VEs and 360-degree films are missing within the selected papers. Most studies have used a randomised controlled trial approach that involves comparison of control (standard education or treatment approach) with experimental XR group. 360-degree filmed scenarios can be experienced on VR HMDs, 3D modelled VEs can be experienced on VR, MR, and AR HMDs. 360-degree filmed scenarios are time and cost-effective when compared with 3D modelled VEs that require more time, money, and advanced computing power for development [39]. 360-degree films can depict real-world scenarios, however, the interactivity is limited or non-existent. Although, 3D modelled VEs can provide greater interactivity, the computational costs involved are high which means photorealism is often sacrificed. Lack of photorealism makes experiences appear clunky [39]

There is continuous advancement in 360-degree filming and production technology, as well as software packages (Unreal Engine, Unity, and Blender) used for the development of 3D modelled VEs. Furthermore, the 3D development and rendering software packages are free to use. Yet, healthcare XR experiences are not at par with gaming experiences due to lack of knowledge, costs, and time constraints. To gain a better understanding of the advantages and drawbacks of using 360-degree filmed and 3D modelled VEs, healthcare XR researchers must widen their research focus.

Additional suggestions for future evaluations

This paper aims to build an understanding of the utilization of the some of the XR HMDs across healthcare pathways, the popularity of HMDs within specific pathways, intervention types, and the evaluation methods used to analyse the effectiveness the HMDs. Based on the findings, some suggestions are made to improve future XR research design and evaluation approaches.

1. Studies within this review tend to focus on evaluating feasibility, acceptability, usability, efficacy, and effectiveness of one type of HMD. XR HMDs offer different experiences in the form of VR, MR and AR. Each type of technology has its pros and cons and will require different approaches for evaluation. In the future, researchers could compare functionalities afforded by different XR HMDs.

2. Most papers present user feedback and/or assessment of healthcare workers and students, or patients. Few papers combine both healthcare providers and patients' responses toward XR intervention, which could be a good step toward improving our knowledge base. XR is still in its early days; development, adoption, scaling up and research priority setting are heavily guided by developers, hospital decision makers and government funding bodies. Including these additional views can present a well-connected picture of the challenges and barriers to XR implementation. This is currently missing from literature.

3. Diversity and inclusion of different samples is missing from the researched literature. XR aims to improve patient outcomes or, patient quality of care because of health worker's improved competency and knowledge retention. Patient perspectives are integral to the development and must include participants representing diversity. This can be achieved through collaborations between different universities and inclusion of multiple hospital sites, specifically ones placed in culturally and economically diverse regions. Issues such as XR accessibility or effect of XR induced motion sickness on different genders can be studied further by including under-represented patient groups.

4. Longitudinal approach toward data collection has not been practiced in any of the studies referenced in this review. XR experiences could be novel and have a one-off effect that cannot be replicated once uniqueness fades or experiences may need to be offered at regular intervals to maintain retention, competence, and distraction.

5. Customised questionnaires are often used within the researched literature. Several studies investigate factors relating to technology perception, process involvement and outcomes through opinions/feedback surveys. This generates qualitative data. Overall usage of validated mixed or quantitative scales is less frequent, especially in user experience and usability evaluations. Validated scales produce reliable, accurate results, and can be easily replicated by others, thereby improving empirical research. Furthermore, user outcomes have not been studied from a behavioural science perspective. The primary aim of XR interventions is to nudge the user toward changing their behaviour. Theoretical models such as Fogg's Behaviour Change Grid, Elaboration Likelihood Model (ELM), Persuasive System Design (PSD) Model, Theory of Planned Behaviour (TPB) and Theory of Reasoned Action (TRA), can be used to design and evaluate XR experiences as behaviour change support systems. To

explore other applicable theoretical models and scales, the recommendation is to research persuasion, user behaviour, neuroscience, and psychology literature. In addition to traditional mixed methods data collection, eye-tracking and brain wave technologies which can be paired with XR HMDs, could provide further insights into user behaviour.

6. XR HMDs can be paired with omni-directional treadmills, haptics gloves and body suits for feedback. The role of these XR HMD add-ons should be studied in more depth.

7. Altspace VR, Horizon World and Engage VR are some examples of social VR learning platforms that are used for education and training. These platforms are easy to operate, often free to use and offer tools for novice users to build virtual worlds (e.g. Horizon World). The implications of these platforms and issues of data privacy must be evaluated within the context of healthcare.

Limitations

This review consists of limited number of studies. Healthcare XR research is a growing field with applications across several different pathways. It is recommended that a systematic review can contribute greatly to the XR evaluation knowledge base. This manuscript does not explore papers through cross tabulation. Future studies can include cross tabulation to establish relationship between variables.

Conclusion

A wide selection of XR HMDs have been used across several different clinical pathways. These interventions are used to improve healthcare workers/students' skills, and patient experiences through pain/anxiety distraction. Various evaluation techniques have been used across literature. This paper categorises the techniques into technology, process, and outcomes evaluations. In the domain of healthcare workers/students, most evaluations are technology or outcome based, qualitative approaches toward assessing usability, feasibility, user experience, and quantitative approaches for measuring competencies and knowledge retention. In the patient's domain, variety of mixed-methods approaches are used to evaluate UX, usability, cybersickness, immersion, presence, distraction, and improved outcomes. Overall, process evaluations are lacking within selected literature. Understanding the process of immersion and engagement afforded by the XR HMD and experience can help in building effective interventions. Future research in the form of systematic reviews is essential to improve evidence based XR evaluation practices.

Supplementary

Appendix A

Appendix A presents HMDs corresponding with papers and authors.

Headset	**Papers**
Samsung Gear VR	[29], [65–71]
HTC Vive	[13], [54], [69], [72–77]
Oculus Rift 1 and 2	[6], [8], [16], [19], [29], [42], [51], [53], [65], [76], [78–90]
Oculus Quest	[69]
Microsoft Hololens	[9], [17], [18], [26], [28], [31], [34], [41], [45–47], [92–106]
Google Glass	[88], [89]
Google Cardboard	[90], [91]
Google Daydream	[92]
Merge VR	[41]
Moverio BT-200	[70], [93], [94]
ODG R-7	[70]
Noon VR	[95]
Voxin VR	[96]
Sony HMZ T1/T2/T3	[37], [97]
Oncomfort	[98]
Moog Simodont	[99]
Xiaozhai V4	[100]
nVisor ST50 Head-Mounted Display for AR	[101]
eMagin Z800 3D	[11], [21], [102], [103]
NJIT RAVR	[104]
VR Box	[105]
Illusion Mask 3D	[103]
Visbox	[106]
NVIS MX90 VR	[107]
HooToo 3D	[74]
Revelation 3D	[108]
AR5DT	[109]
Vuzix iWear VR920	[109], [110]
Sony Glasstron	[111]
Headset not mentioned	[7], [32], [128–131]

Appendix B

Appendix B: Pathways and Interventions

Healthcare Pathway	Healthcare staff and student interventions	Patient interventions
Anatomy Education	1. MR for improving understanding of liver anatomy [117] 2. MR for teaching respiratory gross anatomy [87] 3. Immersive virtual reality as a teaching tool for neuroanatomy [58]	
Cardiology	1. Feasibility of novel visualization techniques: 3D printing (3DP) and augmented reality (AR) in planning transcatheter pulmonary interventions [84] 2. -Feasibility of an augmented reality cardiopulmonary resuscitation training system for health care providers [6] 3. Proof of concept and examine the feasibility and initial user experience of a new MR hologram platform for diagnosis and treatment plan in complex congenital heart disease [86]	1. Reduce pre-operation anxiety [46], [98] 2. Sleep quality in cardiac ICU[95] 3. Teaching patients about the consequences of Arterial Fibrillation and pharmacological stroke prevention [7]

Chiropractic	1. Accuracy and repeatability of projected anatomically correct X-ray images on to a person's skin [75]	
Dental	1. Use of VR within the pre-clinical curriculum in the direct restoration module of the operative dentistry course [99]	1. Pain distraction [10]
Emergency medicine	1. Augmented reality to teach triaging and telemedicine assistance during mass casualty incident [118]	

ENT	1. AR system for accurate, intraoperative localization of pathology and normal anatomic landmarks during open head and neck surgery [83]	1. Smartphone based Virtual Reality Epley Maneuver System (VREMS) for home use [74]
Exercise Science		1. Pain distraction during high intensity biking [51]
Gastroenterology	1. Real world usability analysis of two augmented reality headsets in visceral surgery [30]	1. Pain distraction and procedure satisfaction during colonoscopy [93]

Gerontology		1. Mobility assessment [15]
Gynaecology	1. Train novice surgeons in Africa to perform a virtual radical abdominal (open) hysterectomy (RAH) [63] 2. Virtual Reality Anatomic Model to test Resident Knowledge of Female Pelvic Anatomy [106]	1. Pain distraction during hysterosalpingography [32] 2. Pain distraction during labor [119]
ICU	1. Deliver remote bedside teaching during covid [120]	1. Visual and acoustic stimulation in ICU [25]

Neurology	1. Feasibility of using a low-cost commercially available head-mounted holographic AR device (the Microsoft Hololens) in the operating room when surgeon is fully dressed in PPE [80] 2. Holographic computer for guiding external ventricular drain insertion at the bedside [73]	1. Transfer of motor skills-Sequential Visual Isometric Pinch Task (SVIPT) [20] 2. Training for spatial attention after stroke [69] 3. Patient education before epilepsy surgery and stereotactic implantation of DBS or stereo-EEG electrodes [13] 4. Visual cues taught through hololens to parkinsons patients to improve freezing of gait [14] 5. Tea making task [82] 6. Functional improvements, motivation aspects and clinical effectiveness when using immersive 3D virtual reality versus non-immersive 2D exergaming for patients with parkinsons [52] 7. Effectiveness of immersive VR on upper extremity function in patients with ischemic stroke [116] 8. Intensive virtual reality and robotic based upper limb training compared to usual care, and associated cortical reorganization, in the acute and early sub-acute periods post-stroke [104] 9. Safety and Feasibility of a First-Person View, Full-Body Interaction Game for Telerehabilitation Post-Stroke [60] 10. Embodiment and neuropathic pain caused by spinal cord injury [113] 11. Gaming task procedure maximizes effects of vestibular rehabilitation in unilateral vestibular hypofunction [108] 12. Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease [57] 13. Virtual reality as a tool for evaluation of repetitive rhythmic movements in the elderly and Parkinson's disease patients [110] 14. Pre operative neurosurgical planning[72]

Nursing		1. Pain, fear, and anxiety distraction in children undergoing blood draw [40] 2. Distraction during peripheral intravenous catheter [12], [40], [92], [100], [112]
Opthalmology		1. Visual training for patients with Anisometric Amblyopia [66] 2. Enhanced Depth Navigation Through Augmented Reality Depth Mapping in Patients with Low Vision (Retinitis Pigmentosa) [76]
Orthopaedic	1. Training trainees in VR versus conventional procedures for performing total hip arthroplasty [50] 2. Improve accuracy of Acetabular Cup Orientation in Simulated Total Hip Arthroplasty [81] 3. Improve the surgical skills(intra-articular distal tibial fracture reduction task) education of orthopaedic surgery residents [89]	1. Self-administration of sedation during joint replacement surgery [47] 2. Preoperative virtual reality (VR) experience of 3-dimensional (3D) reconstructed magnetic resonance images (MRIs) on anxiety reduction in patients undergoing arthroscopic knee surgery [44]

Paediatric		1. Immersive virtual reality tour of the operating theater on emergence delirium in children undergoing general anesthesia [121] 2. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children [1]
Pain Management		1. Pain reduction in hospitalized patients [45]
Physiotherapy		1. Effect of VR on body weight support walking [59]

Psychology and Psychiatry		1. Smoke cessation [91] 2. Cognitive Behavioural Therapy (CBT) [37], [97] 3. Psycho-educational tools for depression [68] 4. Repeated Exposure to Perceptual Illusion Challenges Reduces Anxiety Sensitivity Cognitive Concerns [101] 5. Gaze behaviour determination in dyadic interactions [77] 6. Cybersickness during game playing [23] 7. Pain distraction [43], [49], [54], [55], [64], [67], [105], [107], [115] 8. Claustrophobia [61], [102] 9. Fear of heights [48] 10. Social anxiety disorder [21] 11. Small animal phobia [109] 12. Use of VR in patients in Autism Spectrum Disorder [22] 13. Mild cognitive impairment [122]
Radiology	1. Detection of Lung Nodules on Computed Tomography [24] 2. Patient Positioning [65] 3. Guided Lumbar Facet Joint Injections [71]	

Surgery	1. MR visualisation platform to assist intraoperative surgical guidance [85] 2. 360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial [114] 3. VR training tool for orthognathic surgery [62] 4. Effect of using VR surgery on the self-confidence and knowledge of surgical residents performing Le Fort I maxillary osteotomy [4] 5. 360º video for an index-operation (augmented with educational material) [42] 6. The use of head-mounted display eyeglasses for teaching surgical suturing skills [88] 7. Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display to read medical information [70] 8. Training and assessment of digital rectal examination using AR [8]	1. Pain distraction during ambulatory surgery (lipoma resection) [103] 2. Patient pain distraction [11]
Ultrasound	1. Hololens for guiding ultrasound needle placement [94], [123] 2. Clinical application of HMD in sonography [111]	

DIGITAL FUTURES
HUMAN CENTERED DIGITAL INNOVATION

Urology	1. 3D VR models for planning nephrectomy surgery [90] 2. Minimally invasive endoscopic surgery [5] 3. VR for teaching catheter placement [31] 4. 3D mixed reality holograms for preoperative surgical planning of nephron-sparing surgery: evaluation of surgeons' perception [70]	1. 3D VR models for planning nephrectomy surgery [90]
Veterinary	1. Improvement of canine sterilisation surgery skills [96]	

Appendix C

Appendix C: Authors, Domain, Pathways, Outcomes and Evaluation Techniques

Healthcare staff and students skills/outcomes	**Patient skills/outcomes**	**Metrics**
Anatomy Education		
Opinions, usefulness, and usability [117]		i. Users asked to guess lesion location ii. Time taken to complete the task iii. Feedback on Hololens- evaluate the screen, the comfort, and the usefulness HoloLens
Knowledge retention and user experience [87]		1. Knowledge assessment 2. Feedback on user experience
Knowledge retention, and motivation [58]		i. Instructional Materials Motivation Survey (IMMS) ii. Knowledge test
Cardiology		
	Pre-operative anxiety [46]	State Trait Anxiety Inventory (STAI)

	Improved procedure outcome [95]	i. Self report sleep quality measure using Sleep Scale A and the activity tracker FitBit Charge 2.
	Improved procedure outcome, anxiety and pain distraction, immersion and presence [98]	i. patients' subjective opinions, nurses' opinions, applicability, and usability ii. demographic factors iii. Visual Analogue Scale (VAS) iv. Dissociative Experience Scale (DES) 28-items v. Immersion and presence (VAS) vi. Time perception vii. Physiological outcomes
	Knowledge retention [7]	i. Demographics ii. Knowledge and consequence questions (feedback)
Opinions, usability, and effectiveness [84]		i. Opinions- physicians accuracy ii. Usability iii. Potential effectiveness
Competency and satisfaction [6]		i. Technical assessment using AR system ii. Demographics and feedback on satisfaction
Competency and user experience [86]		i. Diagnostic and quality rating questionnaire ii. Technical assessment
Chiropractic		
Accuracy [75]		i. System accuracy and repeatability (technical analysis of system)
Dental		
Competency [99]		i. Competency evaluated by three assessors

	Pain distraction, user experience, nausea and fun [10]	i. Graphic Rating Scale (GRS) ii. Interview with dentist about patient's experience
Emergency Medicine		
Competency [118]		i. Number of triage ii. Duration of triage iii. Quality of triage
ENT		
Face validity and usefulness [83]		i. Customised face validity questionnaire
	Competency and face validity [74]	i. Two expert Otologists reviewed the videos, assigning each participant a score (out of 10) for performance on each step ii. NASA Task Load score
Exercise Science		
	Improved procedure outcome and pain distraction[51]	i. Pain rating feedback
Gastroenterology		
Usability [30]		i. System Usability Scale (SUS)

	Pain and anxiety distraction, improved procedure outcome, and duration[93]	i. Visual Analogue Scale (VAS) ii. Task assessment of the treatment provided
General Medicine		
Knowledge retention, competency, usefulness and usability [8]		i. Technical skills assessment on benchtop model with AR ii. Questionnaire on usability, usefulness
Gerontology		
	Improved procedure outcome, risk assessments, and confidence [15]	i. Technical assessment ii. Physiological Profile Assessment (PPA) iii. The Montreal Cognitive Assessment (MoCA) iv. Activity-specific Balance Confidence (ABC) scale
Gynaecology		
	Anxiety and pain distraction, satisfaction, and immersion [32]	i. Visual Analog Scale (VAS). ii. Subjective responses for affective pain, cognitive pain, and anxiety during procedure; worst pain within 15 min after; patient satisfaction and acceptance with pain management; physiological parameters; adverse effects iii. Technical results of procedure iv. Immersion perception score of the VR system
Knowledge retention and competency [63]		i. predictors of movement and time efficiency

	Pain and anxiety distraction, nausea, and perception [119]	i. Observation of participants ii. Custom questionnaires about pain, anxiety, nausea and perception of VR
Knowledge retention and technology assessment [106]		i. Task assessment ii. Knowledge tests pre and post iii. Technology assessment questionnaire
ICU		
Opinions, acceptability, feasibility, risk, and usefulness [120]	Opinions[120]	i. Customised questionnaire ii. Verbal feedback from patients
	Usability, sickness, immersion, presence, realism , involvement, disorientation, and oculomotor problems[25]	i. Customised questionnaire addressing usability, sickness, oculomotor problems, disorientation, immersion, presence, realism, involvement
Neurology		
	Simulator sickness[20]	i. Simulator sickness questionnaire, adapted from Kennedy, Lane, Berbaum, & Lilienthal (1993)(Kennedy *et al.*, 1993) pre-post ii. Witmer and Singer (1994) presence questionnaire iii. Demographic background questions

	Improve procedure outcome, competency, knowledge retention, cybersickness, engagement, motivation and presence[69]	i. The Dutch version of the Oxford Cognitive Screen (OCS-NL) ii. Language impairment tests iii. Numeric cognition test iv. Verbal and episodic memory tests v. Praxis assessment vi. Executive function assessment vii. Visual confrontation test. viii. Behavioural Inattention Test (BIT) letter cancelation task and figure copy task ix. Simulator Sickness Questionnaire (SSQ) x. Narrative Engagement Scale xi. Intrinsic Motivation Inventory xii. International Television Commission Sense of Presence Inventory (ITC-SOPI)
	Knowledge retention, safety, anxiety distraction and user experience[13]	i. Custom questionnaires to evaluate- knowledge retention, safety, anxiety regarding surgery and feedback
	Improved procedure outcome, mental state, and anxiety and feat distraction[14]	i. FOG (freezing of gait) episodes were annotated by two independent raters from video recordings. ii. Motion data iii. New Freezing of Gait Questionnaire (NFOG-Q) iv. Unified Parkinsons Disease Rating Scale part III (MDS-UPDRS) v. Mini Mental State Examination(MMSE) vi. Fear Avoidance Beliefs questionnaire (FAB)
	Usability, feasibility and acceptability [82]	i. Patients suffering from Alzheimer's disease were tested and post-hoc semi-structured interviews were conducted to assess usability. ii. Studied usability through observation and interviewed for feasibility and acceptability

	Competency and motivation[52]	i. Technical skills assessment ii. Intrinsic Motivation Inventory (IMI)
Feasibility [80]		i. Voice recognition was evaluated against background noise ii. Task assessment
	Improved procedure outcome[116]	i. Fugl-Meyer Upper Extremity (FMUE) assessment. ii. Action Research Arm Test (ARAT) iii. Functional Independence Measure (FIM) iv. Performance Assessment of Self-Care Skills (PASS-IADL and PASS-BADL)
Competency [72]		i. Accuracy of the Hololens localization using neuro-navigation as the gold standard
	Improved procedure outcome[104]	i. Upper Extremity Fugl-Meyer Assessment (UEFMA) ii. Wrist AROM iii. Maximum Pinch Force iv. Wolf Motor Function Test (WMFT) v. Transcraniel Magnetic Stimulation (TMS) mapping
	Safety and usability [60]	i. Demographics ii. Semi structured interview about safety and usability
Technical assessment, feasibility, accuracy, and competency [73]		i. Technical skills assessment

	Leg ownership, pain distraction, and improved procedure outcome[113]	i. Leg Ownership questionnaire ii. Visual Analogue Scale (VAS) iii. Full Body Illusion questionnaire about ownership of virtual body
	Improved procedure outcome, dizziness , anxiety distraction and sickness[108]	i. Task assessment ii. Italian Dizziness Handicap Inventory iii. Activities Specific Balance Confidence scale iv. Zung Instrument for Anxiety Disorders v. Dynamic Gait Index vi. Simulator Sickness Questionnaire (SSQ)
	Sickness, stress, and improved procedure outcome[57]	i. Mini-BESTest (Mini Balance Evaluations Systems Test) for safety ii. Task assessment iii. Simulator Sickness Questionnaire (SSQ) iv. Technical assessment v. Stress Arousal Checklist (SAC)
	Improved procedure outcome and competency [110]	i. Technical assessment for validity and reliability
Nursing		
	Anxiety, fear and pain distraction and fear, and improved procedure outcome [40]	i. Child Fear Scale ii. Children's Anxiety Meter iii. After the blood draw, level of pain experienced was assessed using the Wong–Baker Faces Pain Rating Scale and the fear and anxiety levels experienced by the children during the blood draw were re-evaluated.
	Pain distraction [12]	i. Sociodemographic characteristics of the patients ii. Visual Analog Scale (VAS)

	Pain and fear distraction, and duration of procedure [100]	i. Wong–Baker Faces Pain Rating Scale (WBFPS) ii. Children's Fear Scale (CFS) iii. The time required for successful intravenous insertion
	Improved procedure outcome and pain distraction[124]	i. Demographics ii. Visual Analogue Scale (VAS) iii. Wong-Baker Faces Pain Rating Scale (WBFPS)
	Pain distraction and improved procedure outcome[92]	i. Child-rated Faces Pain Scale-Revised (FPS-R) ii. Caregiver’s rating of child’s distress using Visual Analogue Scale (VAS) iii. Procedural data iv. Feedback from children and caregivers
	Pain distraction[125]	i. Phlebotomist’s feedback on patient pain ii. Parent’s feedback iii. Wong-Baker FACES scale
Ophthalmology		
	Improved procedure outcome[66]	i. Best corrected visual acuity (BCVA)
	Improved procedure outcome[76]	i. Technical assessment of the system
Orthopaedic		

	Improved procedure outcome, user experience, usability, satisfaction[47]	i. Intra-operative propofol use ii. Pattern of propofol over each hour, the amount of adjuvant midazolam or fentanyl used before the case, the overall unmet propofol demand, and postoperative patient satisfaction scores iii. Patient experience - Quality of Recovery Survey (QoR-40) The QoR-40 includes questions regarding patient comfort, emotional state, symptoms and pain iv. Patient feedback about the system
Competency [50]		i. Technical skills assessment ii. Task-specific checklist, error in acetabular component orientation, and procedure duration
	Pain and anxiety distraction, and knowledge retention[44]	i. Amsterdam Preoperative Anxiety and Information Scale score to measure level of anxiety and the need for information in patients undergoing arthroscopic knee surgery. ii. Visual Analogue Scale (VAS)
Competency, face validity, acceptability, applicability [81]		i. Technical skills assessment ii. Perceptions of AR questionnaire iii. Applicability of AR for surgical training
Competency and effectiveness [89]		i. Objective Structured Assessment of Technical Skills checklist (OSATS) Global Rating Scale (GRS) ii. Feedback
Pain Management		

	Pain, satisfaction and user experience[45]	i. Numeric Rating Scale (NRS) ii. Feedback on satisfaction with audio-visual experiences iii. Hospital consumer assessment of Healthcare Providers and Systems (HCAHPS)
Paediatric		
	Incidence and degree of condition, anxiety distraction and post operative behaviour disturbances [121]	i. Observation during stay ii. Paediatric Anaesthesia Emergence Delirium (PAED) scale iii. Modified Yale Preoperative Anxiety Scale -Korean version of mYPAS iv. Postoperative behaviour disturbances
	Anxiety distraction[126]	i. Modified Yale Preoperative Anxiety Scale (mYPAS) ii. Self reported anxiety, pain, delirium, need for analgesia and parental anxiety
Physiotherapy		
	Improved procedure outcome [59]	i. Technical task assessment
Psychology and Psychiatry		

	Improved procedure outcome [91]	i. Baseline-demographics, smoking behaviour, health assessment ii. Post-test assessment-Primary outcome was self-reported abstinence iii. Secondary outcomes included sustained abstinence at 90-day follow-up, adherence to the program, and readiness to quit.
	Improved procedure outcome, immersion [67]	i. Pre-test- Personal characteristics and prior gaming experience, creativity questionnaire about creativity, visualisation, and dreaming ii. Post test-After each VR condition, subjects reported pain score during the electrical stimuli Numeric Rating Scale (NRS). Additional questions about perceived level of immersion in the VR experience
	Cost effectiveness, patient experience[37]	i. Social participation- Social participation (self assessment) tracked using electronic device. ii. Green et al. Paranoid Thoughts Scale (GPTS) iii. Resource Use and Costing- Societal costs were computed by adding (A) the direct medical costs of health care services use including the costs of antipsychotic medication and, in the experimental condition, the additional costs of adjunctive VR-CBT treatment; (B) direct nonmedical costs of travel; and (C) indirect costs stemming from lower productivity. For each participant, cost data over the last 3 months were collected at each of three measurement points. Resource use data, for costing, were collected using the Trimbos Institute and Institute of Medical Technology Assessment Questionnaire for Costs Associated with Psychiatric Illness (TiC-P) iv. Virtual Reality Costs- VR therapy hardware, software and training costs v. Travel Costs- Travel costs arose when participants had to make return trips for receiving health care at health services. vi. Productivity Costs-changes in the participants' work status at baseline and at 3 and 6 months postbaseline using the TiC-P.

	User experience, usability, usefulness, satisfaction, suitability[68]	i. Depressive symptoms were measured using the Patient Health Questionnaire-9 (PHQ-9) scale ii. Participants' satisfaction questionnaire- usefulness of the software, user-friendliness, and overall satisfaction
	Anxiety distraction, improved procedure outcome, depression management and user experience [101]	i. Anxiety Sensitivity Index 3 (ASI-3) ii. Acute Dissociation Inventory–Cognitive (ADI-C) iii. Inventory of Depression and Anxiety Symptoms (IDAS).
	Improved procedure outcome [77]	i. Gaze direction analysis
	VR stimuli, presence, perception of reality, interactivity, enjoyment and cybersickness [23]	i. Self assessed stimuli assessment ii. Spatial presence subscale of International Television Commission-Sense of Presence Inventory (ITC-SOPI) iii. Perceived reality using Popova's perceived reality measures iv. Interactivity measured using effectance and control scales v. Enjoyment using Klimmt, Hartmann and Fray's enjoyment questionnaire vi. Simulator Sickness Questionnaire (SSQ)
	Pain distraction and patient experience [43]	i. Customised questionnaire about pain and experience
	Anxiety distraction, mental states (depression, paranoia, quality of life), improved procedure outcome and patient experience[97]	i. Social Interaction Anxiety Scale ii. Paranoid Thought Scales iii. Beck Depression-II Inventory iv. Manchester Short Assessment of Quality of Life. v. Mental states and experiences were measured with a diary technique called experience sampling method (ESM)
	Pain distraction[105]	i. Visual Analogue Scale (VAS)

	Pain distraction, emotional distress/depression, and quality of life[49]	i. Numeric Pain Rating Scale (NPRS) ii. Patient-Reported Outcomes Measurement Information System (PROMIS®) Item Bank v. 1.0–Emotional Distress–Depression iii. World Health Organization Quality of Life Scale Brief Version (WHOQOL-BREF)
	Pain distraction and user experience[64]	i. Customised questionnaire on VR experience ii. Pain rating
	Pain and anxiety distraction [115]	i. Visual Analogue Scale (VAS) ii. State Anxiety Inventory (SAI) iii. Mini Mental State Examination (MMSE)
	Improved procedure outcome, playability and anxiety distraction[61]	i. Spielberger questionnaire ii. Customised playability questionnaire developed using opinions of experts
	Pain distraction, fun and patient experience[107]	i. Graphic Rating Scale (GRS)
	Improved procedure outcome, fear and sickness[48]	i. Heights Interpretation Questionnaire ii. Acrophobia questionnaire (AQ) iii. Improving Access to Psychological Therapies (IAPT) iv. Simulator Sickness Questionnaire (SSQ)
	Anxiety distraction and technology acceptance[102]	i. Heart rate variability (HRV) analysed using ECGs. ii. State-Trait Anxiety Inventory iii. Technology acceptance model (TAM)

	Anxiety and pain distraction, sickness and satisfaction[41]	i. Visual Analogue Scale (VAS) ii. Coloured Analogue Scale (CAS) iii. Wong-Baker FACES Scale-Revised iv. Childhood Anxiety Sensitivity Index v. Child Presence Measure vi. Malaise Scale vii. Satisfaction questionnaire viii. Questionnaire completed by phlebotomists about patient experience
	Social anxiety, phobia, depression, sickness, presence, and costs[21]	i. Liebowitz Social Anxiety Scale- Self Reported (LSAS-SR) ii. Social Phobia Scale (SPS) iii. Social Interaction Anxiety Scale (SIAS) iv. Fear of Negative Evaluation (FNE) v. Beck Depression Inventory (BDI-II) vi. Social Performance Rating Scale (SPRS) vii. Specific Work for Exposure Applied in Therapy (SWEAT) viii. Simulator Sickness Questionnaire (SSQ) ix. Witmer and Singer's Presence Questionnaire (PQ) x. Gatineau Presence Questionnaire (GPQ) xi. Cost Analysis
	Pain distraction, engagement, sickness, and user experience[54]	i. Visual Analogue Scale (VAS) ii. Numerical Rating Scale (NRS) iii. Customised questionnaire about sickness
	Sickness [55]	i. Simulator Sickness Questionnaire (SSQ)
	Phobia, avoidance , anxiety distraction, improved procedure outcome, expectations and satisfaction[109]	i. Behavioral Avoidance Test (BAT) ii. Fear of Spiders Questionnaire (FSQ) iii. Fear and Avoidance Scales iv. Anxiety Diagnostic Interview Schedule IV v. Clinician Severity Scale (CSS) vi. Customised expectations questionnaire

	Immersion, improved procedure outcome, user experience and presence[22]	i. Demographics questionnaire ii. Wechsler Abbreviated Scale of Intelligence (WASI) iii. Independent Television Commission-Sense of Presence Inventory (ITC-SOPI) iv. Participant observation (verbal and nonverbal behaviours)
	Improved procedure outcome, competency[122]	i.The 15-item Korean version of the Geriatric Depression Scale (SGDS-K) was utilized to measure the subjects' depression symptoms ii. The Seoul Neuropsychological Screening Battery, Dementia version (SNSB-D) iii. Korean color-word Stroop test iv. word fluency tests (category and letter fluency) and VR game based assessments of competencies
Radiology		
Knowledge retention and competency [65]		i. Knowledge retention through an examination compared with expert radiographers
Presence, ergonomics, user experience, competency and sickness [24]		i. Customised questionnaires to assess experience and ergonomics ii. Task assessment
Knowledge retention and competency [71]		i. Task assessment by independent reviewer ii. Task assessment by system
Surgery		
Competency, efficiency, usefulness, functionality and usability [85]		i. Custom questionnaire for feedback and usability- mixed reality component assessment, mixed reality functionality assessment, overall assessment of the visualisation platform

	Blood pressure and pain distraction[103]	i. Blood pressure measurements ii. Visual Analogue Scale (VAS)
Competency, face validity, usability, and suitability [70]		i. Text readability ii. Contrast perception iii. NASA Task Load score iv. Frame rate v. System lag
	Pain distraction and improved procedure outcome[11]	i. Visual Analogue Scale (VAS)
Competency and knowledge retention [114]		i. Technical task assessment
Engagement, knowledge retention, communication, and interpersonal skills [42]		i. Assessment using system ii. Knowledge retention questions iii. Feedback on medical knowledge, interpersonal and communication skills
Technical skills: Understanding training needs, efficacy, usability, acceptability, opinions and human error *Non Technical* skills: Situation awareness, decision making, communication & teamwork, and leadership [4]		i. Le Fort I osteotomy specific face and content validity tests ii. Feedback questionnaire about efficacy, usability, and acceptability iii. Questionnaire investigating non-technical skills

Confidence, knowledge retention, situational awareness, decision making, and user experience [62]		i. Demographics questionnaire ii. Perceived self competence (self confidence scale) iii. Feedback
Competency, confidence, and satisfaction[88]		i. Technical assessment ii. Customised questionnaire about confidence and satisfaction
Knowledge retention, competency, usefulness and usability [8]		i. Technical skills assessment on benchtop model with AR ii. Questionnaire on usability, usefulness
Ultrasound		
Knowledge retention, competency, system evaluation [123]		i. Technical skills assessment ii. NASA Task Load index
Technical assessment, competency, and User experience [94]		i. Technical assessment by independent researcher ii. Customised questionnaire about task and HMD
Competency, sickness, preconceptions, and opinions [111]		i. Demographics ii. Questions about participant health iii. Description of sickness symptoms iv. Technical assessment

Urology		
Competency, User Experience [5]		i. Pre test- Participant demographics, clinical experience and previous HMD exposure at baseline ii. Outcomes - Procedural time and Objective Structured Assessment of Technical Skills (OSATS). Primary outcomes were recorded by an external blinded expert endourologist who was trained in using OSATS. This Global Rating Scale (GRS) includes seven specific domains: (A) respect for tissue, (B) time and motion, (C) instrument handling, (D) handling of the endoscope, (E) flow of the procedure, (F) use of assistants, and (G) knowledge of the procedure iii. Post test feedback questionnaires about HoloLens and the symptoms
Knowledge retention, competency, effectiveness, usability [31]		i. Self evaluation- previous knowledge and interest in modern technologies ii. Objective Structured Clinical Examination (OSCE) iii. Self evaluation- students ranked their ability in the given task iv. NASA Task Load index v. Standard System Usability Scale (SUS)
Competency [90]	Duration of stay, patient experience[90]	i. Technical skills assessment ii. Patient stay duration and discharge iii. Patient conversion to radical nephrectomy, conversion to open surgery, margin status, intraoperative complications, postoperative complications, readmissions, and mortality
User experience and usability [79]		i. Opinions on the use and task performed in MR
Veterinary		
Competency [96]		i. Live surgical technique assessment ii. Subjective opinions

References

1. Eijlers R, Utens EMWJ, Staals LM, et al. Systematic Review and Meta-analysis of Virtual Reality in Pediatrics: Effects on Pain and Anxiety. *Anesth Analg* 2019;129(5):1344-1353. doi:10.1213/ANE.0000000000004165

2. Condon C, Lam WT, Mosley C, Gough S. A systematic review and meta-analysis of the effectiveness of virtual reality as an exercise intervention for individuals with a respiratory condition. *Advances in Simulation* 2020;5(1):33. doi:10.1186/s41077-020-00151-z

3. Radianti J, Majchrzak TA, Fromm J, Wohlgenannt I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Comput Educ* 2020;147:103778. doi:10.1016/j.compedu.2019.103778

4. Pulijala Y, Ma M, Pears M, Peebles D, Ayoub A. Effectiveness of Immersive Virtual Reality in Surgical Training-A Randomized Control Trial. *J Oral Maxillofac Surg* 2018;76(5):1065-1072. doi:10.1016/j.joms.2017.10.002

5. al Janabi HF, Aydin A, Palaneer S, et al. Effectiveness of the HoloLens mixed-reality headset in minimally invasive surgery: a simulation-based feasibility study. *Surg Endosc* 2020;34(3):1143-1149. doi:10.1007/s00464-019-06862-3

6. Balian S, McGovern SK, Abella BS, Blewer AL, Leary M. Feasibility of an augmented reality cardiopulmonary resuscitation training system for health care providers. *Heliyon* 2019;5(8). doi:10.1016/j.heliyon.2019.e02205

7. Balsam P, Borodzicz S, Malesa K, et al. OCULUS study: Virtual reality-based education in daily clinical practice. *Cardiol J* 2019;26(3):260-264. doi:10.5603/CJ.a2017.0154

8. Muangpoon T, Osgouei RH, Escobar-Castillejos D, Kontovounisios C, Bello F. Augmented Reality System for Digital Rectal Examination Training and Assessment: System Validation. *J Med Internet Res* 2020;22(8):e18637. doi:10.2196/18637

9. Pedersen P, Palm H, Ringsted C, Konge L. Virtual-reality simulation to assess performance in hip fracture surgery. *Acta Orthop* 2014;85(4):403-407. doi:10.3109/17453674.2014.917502

10. Atzori B, Lauro Grotto R, Giugni A, Calabrò M, Alhalabi W, Hoffman HG. Virtual Reality Analgesia for Pediatric Dental Patients. *Front Psychol* 2018;9. doi:10.3389/fpsyg.2018.02265

11. Ding J, He Y, Chen L, et al. Virtual reality distraction decreases pain during daily dressing changes following haemorrhoid surgery. *J Int Med Res* 2019;47(9):4380-4388. doi:10.1177/0300060519857862

12. Basak T, Duman S, Demirtas A. Distraction-based relief of pain associated with peripheral intravenous catheterisation in adults: a randomised controlled trial. *J Clin Nurs* 2020;29(5-6):770-777. doi:10.1111/jocn.15131

13. House PM, Pelzl S, Furrer S, et al. Use of the mixed reality tool "VSI Patient Education" for more comprehensible and imaginable patient educations before epilepsy surgery and stereotactic implantation of DBS or stereo-EEG electrodes. *Epilepsy Res* 2020;159:106247. doi:10.1016/j.eplepsyres.2019.106247

14. Janssen S, de Ruyter van Steveninck J, Salim HS, et al. The Effects of Augmented Reality Visual Cues on Turning in Place in Parkinson's Disease Patients With Freezing of Gait. *Front Neurol* 2020;11. doi:10.3389/fneur.2020.00185

15. Sun R, Aldunate RG, Sosnoff JJ. The Validity of a Mixed Reality-Based Automated Functional Mobility Assessment. *Sensors (Basel)* 2019;19(9). doi:10.3390/s19092183

16. Logeswaran A, Munsch C, Chong YJ, Ralph N, McCrossnan J. The role of extended reality technology in healthcare education: Towards a learner-centred approach. *Future Healthc J* 2021;8(1):e79-e84. doi:10.7861/fhj.2020-0112

17. Witmer BG, Singer MJ. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 1998;7(3):225-240. doi:10.1162/105474698565686

18. Slater M, Linakis V, Usoh M, Kooper R. Immersion, presence and performance in virtual environments: an experiment with tri-dimensional chess. In: VRST '96 Association for Computing Machinery; 1996:163-172. doi:10.1145/3304181.3304216

19. Slater M, Sanchez-Vives M. Enhancing our lives with immersive virtual reality. *Frontiers Robotics AI* 2016;3(DEC):74. doi: 10.3389/frobt.2016.00074

20. Juliano JM, Liew SL. Transfer of motor skill between virtual reality viewed using a head-mounted display and conventional screen environments. *J Neuroeng Rehabil* 2020;17(1):48. doi:10.1186/s12984-020-00678-2

21. Bouchard S, Dumoulin S, Robillard G, et al. Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. *Br J Psychiatry* 2017;210(4):276-283. doi:10.1192/bjp.bp.116.184234

22. Newbutt N, Sung C, Kuo HJ, Leahy MJ, Lin CC, Tong B. Brief Report: A Pilot Study of the Use of a Virtual Reality Headset in Autism Populations. *J Autism Dev Disord* 2016;46(9):3166-3176. doi:10.1007/s10803-016-2830-5

23. Shafer DM, Carbonara CP, Korpi MF. Factors Affecting Enjoyment of Virtual Reality Games: A Comparison Involving Consumer-Grade Virtual Reality Technology. *Games Health J* 2019;8(1):15-23. doi:10.1089/g4h.2017.0190

24. Nguyen BJ, Khurana A, Bagley B, et al. Evaluation of Virtual Reality for Detection of Lung Nodules on Computed Tomography. *Tomography* 2018;4(4):204-208. doi:10.18383/j.tom.2018.00053

25. Gerber SM, Jeitziner MM, Wyss P, et al. Visuo-acoustic stimulation that helps you to relax: A virtual reality setup for patients in the intensive care unit. *Sci Rep* 2017;7(1):13228. doi:10.1038/s41598-017-13153-1

26. Lessiter J, Freeman J, Keogh E, Davidoff J. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence* 2001;10:282-297. doi:10.1162/105474601300343612

27. Sutcliffe A, Gault B. Heuristic evaluation of virtual reality applications. *Interact Comput* 2004;16:831-849. doi:10.1016/j.intcom.2004.05.001

28. Tcha-Tokey K, Christmann O, Loup-Escande E, Loup G, Richir S. Towards a Model of User Experience in Immersive Virtual Environments. *Advances in Human-Computer Interaction* 2018;2018:e7827286. doi:10.1155/2018/7827286

29. Brooke J. SUS: A quick and dirty usability scale. Usability Evaluation In Industry. 1st Edition. Boca Raton, Florida, USA: CRC Press; 1995;189. ISBN: 9780429157011

30. Moosburner S, Remde C, Tang P, et al. Real world usability analysis of two augmented reality headsets in visceral surgery. *Artif Organs* 2019;43(7):694-698. doi:10.1111/aor.13396

31. Schoeb DS, Schwarz J, Hein S, et al. Mixed reality for teaching catheter placement to medical students: a randomized single-blinded, prospective trial. *BMC Med Educ* 2020;20(1):510. doi:10.1186/s12909-020-02450-5

32. Wang W, Cheng J, Guo JLC. Usability of Virtual Reality Application Through the Lens of the User Community: A Case Study. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* 2019:1-6. doi:10.1145/3290607.3312816

33. Sutcliffe A, Kaur K. Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology - Behaviour & IT* 2000;19:415-426. doi:10.1080/014492900750052679

34. Joibi J, Kim Y, Shin GW, Lee KJ, Yun M. 1H3-3 A Systematic Review of Extended Reality (XR)'s Head-Mounted Display (HMD) Hardware Design Factors from the Perspective of Usability. *The Japanese Journal of Ergonomics* 2019;55:1H3-3. doi:10.5100/jje.55.1H3-3

35. Beams R, Brown E, Cheng WC, et al. Evaluation Challenges for the Application of Extended Reality Devices in Medicine. *Journal of Digital Imaging 2022;2022*:1-10. doi:10.1007/S10278-022-00622-X

36. XR Health UK. The Growing Value Of XR In Healthcare In United Kingdom 2021. https://www.xrhealthuk.org/the-growing-value-of-xr-in-healthcare [accessed May 11, 2021].

37. Pot-Kolder R, Veling W, Geraets C, et al. Cost-Effectiveness of Virtual Reality Cognitive Behavioral Therapy for Psychosis: Health-Economic Evaluation Within a Randomized Controlled Trial. *J Med Internet Res* 2020;22(5):e17098. doi:10.2196/17098

38. Delshad SD, Almario C v, Fuller G, Luong D, Spiegel BMR. Economic analysis of implementing virtual reality therapy for pain among hospitalized patients. *NPJ Digit Med* 2018;1(1):1-8. doi:10.1038/s41746-018-0026-4

39. Ritter KA, Chambers TL. Three-dimensional modeled environments versus 360 degree panoramas for mobile virtual reality training. *Virtual Real* 2022;26(2):571-581. doi:10.1007/s10055-021-00502-9

40. Özalp Gerçeker G, Ayar D, Özdemir EZ, Bektaş M. Effects of virtual reality on pain, fear and anxiety during blood draw in children aged 5-12 years old: A randomised controlled study. *J Clin Nurs* 2020;29(7-8):1151-1161. doi:10.1111/jocn.15173

41. Gold JI, Mahrer NE. Is Virtual Reality Ready for Prime Time in the Medical Space? A Randomized Control Trial of Pediatric Virtual Reality for Acute Procedural Pain Management. *J Pediatr Psychol* 2018;43(3):266-275. doi:10.1093/jpepsy/jsx129

42. Harrington CM, Kavanagh DO, Wright Ballester G, et al. 360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention. *J Surg Educ* 2018;75(4):993-1000. doi:10.1016/j.jsurg.2017.10.010

43. Hirt PA, Lev-Tov H. The use of virtual reality for bedside procedures. *Br J Dermatol* 2019;181(2):393-394. doi:10.1111/bjd.17682

44. Yang JH, Ryu JJ, Nam E, Lee HS, Lee JK. Effects of Preoperative Virtual Reality Magnetic Resonance Imaging on Preoperative Anxiety in Patients Undergoing Arthroscopic Knee Surgery: A Randomized Controlled Study. *Arthroscopy* 2019;35(8):2394-2399. doi:10.1016/j.arthro.2019.02.037

45. Spiegel B, Fuller G, Lopez M, et al. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. *PLoS One* 2019;14(8):e0219115. doi:10.1371/journal.pone.0219115

46. Hendricks TM, Gutierrez CN, Stulak JM, Dearani JA, Miller JD. The Use of Virtual Reality to Reduce Preoperative Anxiety in First-Time Sternotomy Patients: A Randomized Controlled Pilot Trial. *Mayo Clin Proc* 2020;95(6):1148-1157. doi:10.1016/j.mayocp.2020.02.032

47. Huang MY, Scharf S, Chan PY. Effects of immersive virtual reality therapy on intravenous patient-controlled sedation during orthopaedic surgery under regional anesthesia: A randomized controlled trial. *PLoS One* 2020;15(2):e0229320. doi:10.1371/journal.pone.0229320

48. Freeman D, Haselton P, Freeman J, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *Lancet Psychiatry* 2018;5(8):625-632. doi:10.1016/S2215-0366(18)30226-8

49. Benham S, Kang M, Grampurohit N. Immersive Virtual Reality for the Management of Pain in Community-Dwelling Older Adults. *OTJR (Thorofare N J)* 2019;39(2):90-96. doi:10.1177/1539449218817291

50. Logishetty K, Rudran B, Cobb JP. Virtual reality training improves trainee performance in total hip arthroplasty: a randomized controlled trial. *Bone Joint J* 2019;101-B(12):1585-1592. doi:10.1302/0301-620X.101B12.BJJ-2019-0643.R1

51. Wender CLA, Ahn SJ, O'Connor PJ. Interactive Virtual Reality Reduces Quadriceps Pain during High-Intensity Cycling. *Med Sci Sports Exerc* 2019;51(10):2088-2097. doi:10.1249/MSS.0000000000002016

52. Cikajlo I, Peterlin Potisk K. Advantages of using 3D virtual reality based training in persons with Parkinson's disease: a parallel study. *J Neuroeng Rehabil* 2019;16(1):119. doi:10.1186/s12984-019-0601-1

53. Wang YL, Gao HX, Wang JS, et al. Immersive virtual reality as analgesia for women during hysterosalpingography: study protocol for a randomized controlled trial. *Trials* 2020;21(1):102. doi:10.1186/s13063-019-4023-y

54. Jones T, Moore T, Choo J. The Impact of Virtual Reality on Chronic Pain. *PLoS One* 2016;11(12). doi:10.1371/journal.pone.0167523

55. Tong X, Gromala D, Gupta D, Squire P. Usability Comparisons of Head-Mounted vs. Stereoscopic Desktop Displays in a Virtual Reality Environment with Pain Patients. *Stud Health Technol Inform* 2016;220:424-431. http://www.ncbi.nlm.nih.gov/pubmed/27046617

56. Rahm S, Wieser K, Wicki I, Holenstein L, Fucentese SF, Gerber C. Performance of medical students on a virtual reality simulator for knee arthroscopy: an analysis of learning curves and predictors of performance. *BMC Surg* 2016;16(1):14. doi:10.1186/s12893-016-0129-2

57. Kim A, Darakjian N, Finley JM. Walking in fully immersive virtual environments: an evaluation of potential adverse effects in older adults and individuals with Parkinson's disease. *J Neuroeng Rehabil* 2017;14. doi:10.1186/s12984-017-0225-2

58. Stepan K, Zeiger J, Hanchuk S, et al. Immersive virtual reality as a teaching tool for neuroanatomy. *Int Forum Allergy Rhinol* 2017;7(10):1006-1013. doi:10.1002/alr.21986

59. Odin A, Faletto-Passy D, Assaban F, Pérennou D. Modulating the internal model of verticality by virtual reality and body-weight support walking: A pilot study. *Ann Phys Rehabil Med* 2018;61(5):292-299. doi:10.1016/j.rehab.2018.07.003

60. Proffitt R, Warren J, Lange B, Chang CY. Safety and Feasibility of a First-Person View, Full-Body Interaction Game for Telerehabilitation Post-Stroke. *Int J Telerehabil* 2018;10(1):29-36. doi:10.5195/ijt.2018.6250

61. Rahani VK, Vard A, Najafi M. Claustrophobia Game: Design and Development of a New Virtual Reality Game for Treatment of Claustrophobia. *J Med Signals Sens* 2018;8(4):231-237. doi:10.4103/jmss.JMSS_27_18

62. Pulijala Y, Ma M, Pears M, Peebles D, Ayoub A. An innovative virtual reality training tool for orthognathic surgery. *Int J Oral Maxillofac Surg* 2018;47(9):1199-1205. doi:10.1016/j.ijom.2018.01.005

63. Bing EG, Parham GP, Cuevas A, et al. Using Low-Cost Virtual Reality Simulation to Build Surgical Capacity for Cervical Cancer Treatment. *J Glob Oncol* 2019;5. doi:10.1200/JGO.18.00263

64. Matamala-Gomez M, Diaz Gonzalez AM, Slater M, Sanchez-Vives M v. Decreasing Pain Ratings in Chronic Arm Pain Through Changing a Virtual Body: Different Strategies for Different Pain Types. *J Pain* 2019;20(6):685-697. doi:10.1016/j.jpain.2018.12.001

65. Sapkaroski D, Mundy M, Dimmock MR. Immersive virtual reality simulated learning environment versus role-play for empathic clinical communication training. *J Med Radiat Sci* 2022;69(1):56-65. doi:10.1002/JMRS.555

66. Halička J, Sahatqija E, Krasňanský M, Kapitánová K, Fedorová M, Žiak P. Visual Training in Virtual Reality in Adult Patients with Anisometric Amblyopia. *Cesk Slov Oftalmol* 2020;76(1):24-28. doi:10.31348/2020/3

67. Lier EJ, Oosterman JM, Assmann R, de Vries M, van Goor H. The effect of Virtual Reality on evoked potentials following painful electrical stimuli and subjective pain. *Sci Rep* 2020;10(1):9067. doi:10.1038/s41598-020-66035-4

68. Migoya-Borja M, Delgado-Gómez D, Carmona-Camacho R, et al. Feasibility of a Virtual Reality-Based Psychoeducational Tool (VRight) for Depressive Patients :7. https://www.liebertpub.com/doi/pdf/10.1089/cyber.2019.0497

69. Huygelier H, Schraepen B, Lafosse C, et al. An immersive virtual reality game to train spatial attention orientation after stroke: A feasibility study. *Appl Neuropsychol Adult* 2020:1-21. doi:10.1080/23279095.2020.1821030

70. Qian L, Barthel A, Johnson A, et al. Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display. *Int J Comput Assist Radiol Surg* 2017;12(6):901-910. doi:10.1007/s11548-017-1564-y

71. Agten CA, Dennler C, Rosskopf AB, Jaberg L, Pfirrmann CWA, Farshad M. Augmented Reality-Guided Lumbar Facet Joint Injections. *Invest Radiol* 2018;53(8):495-498. doi:10.1097/RLI.0000000000000478

72. Incekara F, Smits M, Dirven C, Vincent A. Clinical Feasibility of a Wearable Mixed-Reality Device in Neurosurgery. *World Neurosurg* 2018;118:e422-e427. doi:10.1016/j.wneu.2018.06.208

73. Li Y, Chen X, Wang N, et al. A wearable mixed-reality holographic computer for guiding external ventricular drain insertion at the bedside. *J Neurosurg* 2018;131(5):1599-1606. doi:10.3171/2018.4.JNS18124

74. Tabanfar R, Chan HHL, Lin V, Le T, Irish JC. Development and face validation of a Virtual Reality Epley Maneuver System (VREMS) for home Epley treatment of benign paroxysmal positional vertigo: A randomized, controlled trial. *Am J Otolaryngol* 2018;39(2):184-191. doi:10.1016/j.amjoto.2017.11.006

75. Aaskov J, Kawchuk GN, Hamaluik KD, Boulanger P, Hartvigsen J. X-ray vision: the accuracy and repeatability of a technology that allows clinicians to see spinal X-rays superimposed on a person's back. *PeerJ* 2019;7:e6333. doi:10.7717/peerj.6333

76. Angelopoulos AN, Ameri H, Mitra D, Humayun M. Enhanced Depth Navigation Through Augmented Reality Depth Mapping in Patients with Low Vision. *Sci Rep* 2019;9(1):11230. doi:10.1038/s41598-019-47397-w

77. Binetti N, Cheng T, Mareschal I, Brumby D, Julier S, Bianchi-Berthouze N. Assumptions about the positioning of virtual stimuli affect gaze direction estimates during Augmented Reality based interactions. *Sci Rep* 2019;9(1):2566. doi:10.1038/s41598-019-39311-1

78. Skulmowski A, Pradel S, Kühnert T, Brunnett G, Rey GD. Embodied learning using a tangible user interface: The effects of haptic perception and selective pointing on a spatial learning task. *Comput Educ* 2016;92-93:64-75. doi:10.1016/j.compedu.2015.10.011

79. Checcucci E, Amparore D, Pecoraro A, et al. 3D mixed reality holograms for preoperative surgical planning of nephron-sparing surgery: evaluation of surgeons' perception. *Minerva Urol Nefrol* 2019. doi:10.23736/S0393-2249.19.03610-5

80. Kubben P, Sinlae R. Feasibility of using a low-cost head-mounted augmented reality device in the operating room. *Surg Neurol Int* 2019;10:26. doi:10.4103/sni.sni_228_18

81. Logishetty K, Western L, Morgan R, Iranpour F, Cobb JP, Auvinet E. Can an Augmented Reality Headset Improve Accuracy of Acetabular Cup Orientation in Simulated THA? A Randomized Trial. *Clin Orthop Relat Res* 2019;477(5):1190-1199. doi:10.1097/CORR.0000000000000542

82. Rohrbach N, Gulde P, Armstrong AR, et al. An augmented reality approach for ADL support in Alzheimer's disease: a crossover trial. *J Neuroeng Rehabil* 2019;16(1):66. doi:10.1186/s12984-019-0530-z

83. Rose AS, Kim H, Fuchs H, Frahm JM. Development of augmented-reality applications in otolaryngology–head and neck surgery. *Laryngoscope* 2019;129(S3):S1-S11. doi:10.1002/lary.28098

84. Witowski J, Darocha S, Kownacki Ł, et al. Augmented reality and three-dimensional printing in percutaneous interventions on pulmonary arteries. *Quant Imaging Med Surg* 2019;9(1):23-29. doi:10.21037/qims.2018.09.08

85. Cartucho J, Shapira D, Ashrafian H, Giannarou S. Multimodal mixed reality visualisation for intraoperative surgical guidance. *Int J Comput Assist Radiol Surg* 2020;15(5):819-826. doi:10.1007/s11548-020-02165-4

86. Brun H, Bugge R a. B, Suther LKR, et al. Mixed reality holograms for heart surgery planning: first user experience in congenital heart disease. *Eur Heart J Cardiovasc Imaging* 2019;20(8):883-888. doi:10.1093/ehjci/jey184

87. Robinson BL, Mitchell TR, Brenseke BM. Evaluating the Use of Mixed Reality to Teach Gross and Microscopic Respiratory Anatomy. *Med Sci Educ* Published online April 21, 2020:1-4. doi:10.1007/s40670-020-01064-2

88. Peden RG, Mercer R, Tatham AJ. The use of head-mounted display eyeglasses for teaching surgical skills: A prospective randomised study. *Int J Surg* 2016;34:169-173. doi:10.1016/j.ijsu.2016.09.002

89. Dickerson P, Grande S, Evans D, Levine B, Coe M. Utilizing Intraprocedural Interactive Video Capture With Google Glass for Immediate Postprocedural Resident Coaching. *J Surg Educ* 2019;76(3):607-619. doi:10.1016/j.jsurg.2018.10.002

90. Shirk JD, Thiel DD, Wallen EM, et al. Effect of 3-Dimensional Virtual Reality Models for Surgical Planning of Robotic-Assisted Partial Nephrectomy on Surgical Outcomes. *JAMA Netw Open* 2019;2(9). doi:10.1001/jamanetworkopen.2019.11598

91. Goldenhersch E, Thrul J, Ungaretti J, Rosencovich N, Waitman C, Ceberio MR. Virtual Reality Smartphone-Based Intervention for Smoking Cessation: Pilot Randomized Controlled Trial on Initial Clinical Efficacy and Adherence. *J Med Internet Res* 2020;22(7):e17571. doi:10.2196/17571

92. Chan E, Hovenden M, Ramage E, et al. Virtual Reality for Pediatric Needle Procedural Pain: Two Randomized Clinical Trials. *J Pediatr* 2019;209:160-167.e4. doi:10.1016/j.jpeds.2019.02.034

93. Umezawa S, Higurashi T, Uchiyama S, et al. Visual distraction alone for the improvement of colonoscopy-related pain and satisfaction. *World journal of gastroenterology : WJG* 2015;21:4707-4714. doi:10.3748/wjg.v21.i15.4707

94. Kaneko N, Tsunoda M, Mitsuhashi M, et al. Ultrasound-Guided Fine-Needle Aspiration in the Neck Region Using an Optical See-Through Head-Mounted Display: A Randomized Controlled Trial. *J Ultrasound Med* 2017;36(10):2071-2077. doi:10.1002/jum.14237

95. Lee SY, Kang J. Effect of virtual reality meditation on sleep quality of intensive care unit patients: A randomised controlled trial. *Intensive Crit Care Nurs* 2020;59:102849. doi:10.1016/j.iccn.2020.102849

96. Hunt JA, Heydenburg M, Anderson SL, Thompson RR. Does virtual reality training improve veterinary students' first canine surgical performance? *Vet Rec* 2020;186(17):562. doi:10.1136/vr.105749

97. Geraets CNW, Veling W, Witlox M, Staring ABP, Matthijssen SJMA, Cath D. Virtual reality-based cognitive behavioural therapy for patients with generalized social anxiety disorder: a pilot study. *Behavioural and Cognitive Psychotherapy* 2019;47(6):745-750. doi:10.1017/S1352465819000225

98. Rousseaux F, Faymonville ME, Nyssen AS, et al. Can hypnosis and virtual reality reduce anxiety, pain and fatigue among patients who undergo cardiac surgery: a randomised controlled trial. *Trials* 2020;21(1):330. doi:10.1186/s13063-020-4222-6

99. Murbay S, Neelakantan P, Chang JWW, Yeung S. Evaluation of the introduction of a dental virtual simulator on the performance of undergraduate dental students in the pre-clinical operative dentistry course. *Eur J Dent Educ* 2020;24(1):5-16. doi:10.1111/eje.12453

100. Chen YJ, Cheng SF, Lee PC, Lai CH, Hou IC, Chen CW. Distraction using virtual reality for children during intravenous injections in an emergency department: A randomised trial. *J Clin Nurs* 2020;29(3-4):503-510. doi:10.1111/jocn.15088

101. Albanese BJ, Bauer BW, Raudales AM, Capron DW, Schmidt NB. Repeated Exposure to Perceptual Illusion Challenges Reduces Anxiety Sensitivity Cognitive Concerns: Evidence From a Randomized Clinical Trial. *Behav Ther* 2019;50(6):1173-1184. doi:10.1016/j.beth.2019.04.003

102. Li YY, Chiu PH, Yeh SC, Zhou C. Effects of Virtual Reality and Augmented Reality on Induced Anxiety. In: *2017 5th International Conference on Enterprise Systems (ES)* ; 2017:132-138. doi:10.1109/ES.2017.28

103. Mosso Vázquez JL, Mosso Lara D, Mosso Lara JL, Miller I, Wiederhold MD, Wiederhold BK. Pain Distraction During Ambulatory Surgery: Virtual Reality and Mobile Devices. *Cyberpsychol Behav Soc Netw* 2019;22(1):15-21. doi:10.1089/cyber.2017.0714

104. Patel J, Fluet G, Qiu Q, et al. Intensive virtual reality and robotic based upper limb training compared to usual care, and associated cortical reorganization, in the acute and early sub-acute periods post-stroke: a feasibility study. *J Neuroeng Rehabil* 2019;16(1):92. doi:10.1186/s12984-019-0563-3

105. Karaman D, Erol F, Yılmaz D, Dikmen Y. Investigation of the effect of the virtual reality application on experimental pain severity in healthy. *Rev Assoc Med Bras (1992)*. 2019;65(3):446-451. doi:10.1590/1806-9282.65.3.446

106. Ellington DR, Shum PC, Dennis EA, Willis HL, Szychowski JM, Richter HE. Female Pelvic Floor Immersive Simulation: A Randomized Trial to Test the Effectiveness of a Virtual Reality Anatomic Model on Resident Knowledge of Female Pelvic Anatomy. *J Minim Invasive Gynecol* 2019;26(5):897-901. doi:10.1016/j.jmig.2018.09.003

107. Soltani M, Drever SA, Hoffman HG, et al. Virtual reality analgesia for burn joint flexibility: A randomized controlled trial. *Rehabil Psychol* 2018;63(4):487-494. doi:10.1037/rep0000239

108. Micarelli A, Viziano A, Augimeri I, Micarelli D, Alessandrini M. Three-dimensional head-mounted gaming task procedure maximizes effects of vestibular rehabilitation in unilateral vestibular hypofunction: a randomized controlled pilot trial. *Int J Rehabil Res* 2017;40(4):325-332. doi:10.1097/MRR.0000000000000244

109. Botella C, Pérez-Ara MÁ, Bretón-López J, Quero S, García-Palacios A, Baños RM. In Vivo versus Augmented Reality Exposure in the Treatment of Small Animal Phobia: A Randomized Controlled Trial. *PLoS One* 2016;11(2):e0148237. doi:10.1371/journal.pone.0148237

110. Arias P, Robles-García V, Sanmartín G, Flores J, Cudeiro J. Virtual Reality as a Tool for Evaluation of Repetitive Rhythmic Movements in the Elderly and Parkinson's Disease Patients. *PLoS One* 2012;7(1):e30021. doi:10.1371/journal.pone.0030021

111. Takeshita H, Kihara K, Yoshida S, et al. Clinical application of a modern high-definition head-mounted display in sonography. *J Ultrasound Med* 2014;33(8):1499-1504. doi:10.7863/ultra.33.8.1499

112. Aydın Aİ, Özyazıcıoğlu N. Using a Virtual Reality Headset to Decrease Pain Felt During a Venipuncture Procedure in Children. *J Perianesth Nurs* 2019;34(6):1215-1221. doi:10.1016/j.jopan.2019.05.134

113. Pozeg P, Palluel E, Ronchi R, et al. Virtual reality improves embodiment and neuropathic pain caused by spinal cord injury. *Neurology* 2017;89(18):1894-1903. doi:10.1212/WNL.0000000000004585

114. Yoganathan S, Finch DA, Parkin E, Pollard J. 360° virtual reality video for the acquisition of knot tying skills: A randomised controlled trial. *Int J Surg* 2018;54(Pt A):24-27. doi:10.1016/j.ijsu.2018.04.002

115. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: A randomized control trial. *Palliat Support Care* 2019;17(1):29-34. doi:10.1017/S1478951518000639

116. Ögün MN, Kurul R, Yaşar MF, Turkoglu SA, Avci Ş, Yildiz N. Effect of Leap Motion-based 3D Immersive Virtual Reality Usage on Upper Extremity Function in Ischemic Stroke Patients. *Arq Neuropsiquiatr* 2019;77(10):681-688. doi:10.1590/0004-282X20190129

117. Pelanis E, Kumar RP, Aghayan DL, et al. Use of mixed reality for improved spatial understanding of liver anatomy. *Minim Invasive Ther Allied Technol* 2020;29(3):154-160. doi:10.1080/13645706.2019.1616558

118. Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M. Technical Support by Smart Glasses During a Mass Casualty Incident: A Randomized Controlled Simulation Trial on Technically Assisted Triage and Telemedical App Use in Disaster Medicine. *J Med Internet Res* 2019;21(1):e11939. doi:10.2196/11939

119. Frey DP, Bauer ME, Bell CL, et al. Virtual Reality Analgesia in Labor: The VRAIL Pilot Study-A Preliminary Randomized Controlled Trial Suggesting Benefit of Immersive Virtual Reality Analgesia in Unmedicated Laboring Women. *Anesth Analg* 2019;128(6):e93-e96. doi:10.1213/ANE.0000000000003649

120. Sivananthan A, Gueroult A, Zijlstra G, et al. Using Mixed Reality Headsets to Deliver Remote Bedside Teaching During the COVID-19 Pandemic: Feasibility Trial of HoloLens 2. JMIR Form Res 2022;6(5):e35674. doi: 10.2196/35674

121. Ryu JH, Oh AY, Yoo HJ, Kim JH, Park JW, Han SH. The effect of an immersive virtual reality tour of the operating theater on emergence delirium in children undergoing general anesthesia: A randomized controlled trial. *Paediatr Anaesth* 2019;29(1):98-105. doi:10.1111/pan.13535

122. Park JH, Liao Y, Kim DR, et al. Feasibility and Tolerability of a Culture-Based Virtual Reality (VR) Training Program in Patients with Mild Cognitive Impairment: A Randomized Controlled Pilot Study. *Int J Environ Res Public Health* 2020;17(9). doi:10.3390/ijerph17093030

123. Rüger C, Feufel MA, Moosburner S, Özbek C, Pratschke J, Sauer IM. Ultrasound in augmented reality: a mixed-methods evaluation of head-mounted displays in image-guided

interventions. *Int J Comput Assist Radiol Surg* 2020;15(11):1895-1905. doi:10.1007/s11548-020-02236-6

124. Aydin A, Muir GH, Graziano ME, Khan MS, Dasgupta P, Ahmed K. Validation of the GreenLight™ Simulator and development of a training curriculum for photoselective vaporisation of the prostate. *BJU Int* 2015;115(6):994-1003. doi:10.1111/bju.12842

125. Gerçeker GÖ, Binay Ş, Bilsin E, Kahraman A, Yılmaz HB. Effects of Virtual Reality and External Cold and Vibration on Pain in 7- to 12-Year-Old Children During Phlebotomy: A Randomized Controlled Trial. *J Perianesth Nurs* 2018;33(6):981-989. doi:10.1016/j.jopan.2017.12.010

126. Eijlers R, Dierckx B, Staals LM, et al. Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: A randomised controlled trial. *Eur J Anaesthesiol* 2019;36(10):728-737. doi:10.1097/EJA.0000000000001059

Abbreviations

Anxiety Sensitivity Index 3 (ASI-3)

Augmented Reality (AR)

Child-rated Faces Pain Scale-Revised (FPS-R)

Coloured Analogue Scale (CAS)

Elaboration Likelihood Model (ELM),

Extended Reality (XR)

Gatineau Presence Questionnaire (GPQ)

Global Rating Scales (GRS)

Graphic Rating Scale (GRS),

Head-Mounted Display (HMD)

Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)

Human Computer Interaction (HCI)

Instructional Materials Motivation Survey (IMMS)

Inventory of Depression and Anxiety Symptoms (IDAS)

Mixed Reality (MR)

Modified Yale Preoperative Anxiety Scale (mYPAS)

NASA Task Load (NASA -TLX)

Numeric Rating Scale, State Trait Anxiety Inventory (STAI),

Objective Structure Assessment of Technical Skills (OSATS)

Objective Structured Clinical Examination (OSCE)

Persuasive System Design (PSD) Model

Quality of Recovery Survey (QoR-40)

Simulator Sickness Questionnaire (SSQ)

System Usability Scale (SUS)

Technology Acceptance Model (TAM)

Theory of Planned Behaviour (TPB)

Theory of Reasoned Action (TRA),

User Experience (UX)

Virtual Environment (VE)

Virtual Reality (VR)

Visual Analogue Scale (VAS)