



UNIVERSIDAD
DE ATACAMA

FACULTAD DE INGENIERÍA
DEPTO. DE ING. INFORMÁTICA Y CIENCIAS DE LA COMPUTACIÓN

**TÉCNICAS DE EVALUACIÓN DE
USABILIDAD PARA ENTORNOS
VIRTUALES: UN ESTUDIO EXPLORATORIO**

Trabajo de título presentado para obtener el título de
Ingeniero Civil en Computación e Informática

Profesor Guía: John W. Castro Llanos

Gianina Amalia Madrigal López

Copiapó, Octubre 2023, Chile.



UNIVERSIDAD
DE ATACAMA

FACULTAD DE INGENIERÍA
DEPTO. DE ING. INFORMÁTICA Y CIENCIAS DE LA COMPUTACIÓN

**TÉCNICAS DE EVALUACIÓN DE
USABILIDAD PARA ENTORNOS
VIRTUALES: UN ESTUDIO EXPLORATORIO**

Trabajo de título presentado para obtener el título de
Ingeniero Civil en Computación e Informática

Profesor Guía:
John W. Castro Llanos

Miembros del Comité:
Vladimir Riffo Bouffanais
Héctor Cornide Reyes

Gianina Amalia Madrigal López

Copiapó, Octubre 2023, Chile.

*Esta tesis se la dedico a mi
mamá y mis seres queridos,
quienes se han encontrado
apoyando mi proceso de
aprendizaje*

Agradecimientos

Quiero expresar mi agradecimiento a todas aquellas personas que han contribuido a la realización de este trabajo. En primer lugar, agradezco a mi madre por su constante apoyo y por inculcarme valores como la perseverancia y la dedicación, sin los cuales no habría logrado llegar hasta aquí. También quiero agradecer a mi mejor amiga Constanza, quien siempre estuvo presente durante los momentos difíciles y me brindó su amistad incondicional. Gracias por tu aliento y tus consejos, sin duda han sido de gran ayuda en todo este proceso. De igual manera, quiero agradecer a mis compañeros de carrera, quienes han sido un pilar fundamental en mi formación académica. Agradezco su apoyo constante y la motivación que me han brindado a lo largo de este camino, así como también el haber compartido conmigo experiencias inolvidables que han dejado una huella en mi vida. Sin su presencia y colaboración, no habría sido posible llegar hasta aquí. Por último, agradezco a mi profesor guía, John W. Catro, por su invaluable orientación, paciencia y conocimientos impartidos a lo largo de todo el proyecto. Gracias por compartir su tiempo y experiencia para hacer posible este trabajo. Todos ustedes han sido una pieza fundamental en este camino que hoy culmina con éxito. Su apoyo y motivación han sido esenciales para alcanzar mis metas. Una vez más, muchas gracias.

Resumen

La realidad extendida es una tecnología que crea experiencias inmersivas en diversos campos como educación, salud y seguridad. Dado el amplio rango de usuarios que interactúan con entornos de realidad extendida, la usabilidad es esencial para mejorar la eficiencia, satisfacción y productividad del usuario. La evaluación de la usabilidad se realiza mediante pruebas que identifican problemas de diseño, como asumir conocimientos que los usuarios no tienen o imponer una carga mental excesiva. Esta evaluación es crucial para evitar la insatisfacción del usuario con el producto final. La investigación sobre la evaluación de la usabilidad en entornos virtuales está fragmentada en diversas bases de datos científicas. Hasta ahora, no existe un trabajo que compile y presente de manera integral las técnicas utilizadas para evaluar la usabilidad en estos sistemas. Esto implica que profesionales e investigadores deban invertir mucho tiempo en buscar, revisar, extraer y sintetizar información relevante de múltiples fuentes. Esta dificultad se acentúa debido a las constantes contribuciones bibliográficas de la comunidad científica. Este trabajo realiza un mapeo sistemático de literatura desde 2018 hasta junio de 2022 en las bases de datos Scopus, *IEEE Xplore* y *Web of Science*, identificando 30 estudios primarios que abordan la evaluación de entornos de realidad extendida. Se destacan tres categorías de entornos: realidad virtual, realidad aumentada y realidad mixta. La revisión revela desafíos en la aplicación de técnicas de evaluación, incluyendo problemas de muestra y limitaciones de tiempo. Se identifican 14 técnicas de evaluación, como entrevistas, evaluación heurística, encuestas, experimentos de usabilidad y entre otras. Algunas técnicas para la evaluación de la usabilidad se han adaptado, como encuestas y recorridos cognitivos para la evaluación de la usabilidad. Combinaciones de técnicas también se emplearon para una evaluación más completa. Además, se han desarrollado *frameworks* para

evaluar la usabilidad en entornos de realidad virtual. Algunos estudios proponen nuevas técnicas, como la evaluación de usabilidad en grupo basada en escenarios y un método automatizado de detección de problemas de usabilidad. Finalmente, se resalta la importancia de la usabilidad en estos entornos para mejorar la experiencia del usuario y se proponen futuras investigaciones en la aplicación de técnicas y estándares del área de la Interacción Persona-Computadora.

Palabras clave: Evaluación de usabilidad, Entornos virtuales, Estudio exploratorio.

Índice general

Índice general	v
Índice de figuras	vii
Índice de tablas	viii
1. Introducción	1
1.1. Objetivos	5
1.1.1. Objetivo General	5
1.1.2. Objetivos Específicos	5
1.1.3. Preguntas	5
1.2. Metodología	6
1.3. Estructura del Trabajo	7
1.4. Publicación Derivada	8
2. Marco Teórico y Trabajos Relacionados	9
2.1. Usabilidad	9
2.2. Evaluación de la Usabilidad	10
2.3. Tipos de Entornos	11
2.3.1. Realidad Virtual	11
2.3.2. Realidad Aumentada	12
2.3.3. Realidad Mixta	12
2.4. Trabajos Relacionados	13
3. Metodología y Síntesis	15
3.1. Identificación de los Estudios Primarios	15
3.1.1. Definición de la Estrategia de Búsqueda	16

3.1.2.	Creación de la Cadena de Búsqueda	17
3.1.3.	Definición de los Criterios de Selección	20
3.1.4.	Selección de los Estudios Primarios	22
3.1.5.	Extraer y Sintetizar la Información	22
3.2.	Síntesis de los Estudios Primarios	24
4.	Análisis de Resultados	27
4.1.	Técnicas de Usabilidad para Evaluar Entornos Virtuales	27
4.2.	Entornos de Realidad Extendida Evaluados	34
4.3.	Principales Problemas y Desafíos en la Aplicación de Técnicas para Evaluar la Usabilidad de Entornos Virtuales	35
4.4.	Cómo se están Aplicando las Técnicas para la Evaluación de la Usa- bilidad de Entornos Virtuales	37
5.	Discusión y Amenazas a la Validez	46
5.1.	Discusión	46
5.2.	Amenazas a la Validez	51
6.	Conclusiones	52
	Referencias	57
	Anexo A - Publicación Derivada	65
	Anexo B - Estudios Primarios	82

Índice de figuras

2.1. Entornos de realidad virtual, (a) HTC Vive, (b) Usuario usando un HTC Vive. Tomado de Tham <i>et al.</i> (2018)	11
2.2. Entornos de realidad aumentada, (a) Magic Book, (b) Studierstube, son sistemas de RA colaborativos, con dos usuarios que llevan PMC transparentes según Azuma <i>et al.</i> (2001)	12
2.3. Entornos de realidad mixta, (a) Ritchie, (b) Humanoide Virtual, son agentes de realidad mixta reportados por Holz <i>et al.</i> (2011)	13
3.1. Tareas realizadas para obtener las palabras claves	18
3.2. Tareas realizadas para conseguir los estudios preseleccionados de las diferentes bases de datos	21
3.3. Tareas realizadas en el proceso de selección de los estudios primarios .	23
3.4. Tareas realizadas para extraer y sintetizar la información de los estudios primarios	24
3.5. Etapas de la metodología a utilizar (EMS)	26
5.1. Aspectos importantes de la investigación	47

Índice de tablas

3.1. Tabla de frecuencia de palabras obtenidas	17
3.2. Cadena de palabras	19
3.3. Campos de búsqueda por base de datos	20
3.4. Número de estudios restantes después de filtrar los resultados de la base de datos	22
4.1. Técnicas de evaluación de usabilidad	34
4.2. Combinación de las técnicas más usadas	41
B.1. Estudios primarios extraídos de la cadena seleccionada	82
B.1. Estudios primarios extraídos de la cadena seleccionada	83
B.1. Estudios primarios extraídos de la cadena seleccionada	84
B.1. Estudios primarios extraídos de la cadena seleccionada	85

Capítulo 1

Introducción

La usabilidad es uno de los indicadores más críticos para determinar la calidad de un producto de software [Castro *et al.*, 2022; Losana *et al.*, 2021; Ren *et al.*, 2022] y corresponde a la forma en que un grupo de usuarios puede utilizar un sistema software para alcanzar objetivos específicos con eficacia, eficiencia y satisfacción [Hertzum, 2020]. La evaluación de la usabilidad se lleva a cabo mediante pruebas de usabilidad ya que, a pesar de los esfuerzos de los desarrolladores de software, los primeros diseños tienden a ser defectuosos ya que pueden, por ejemplo [Hertzum, 2020]: (i) Asumir conocimientos que los usuarios no poseen, (ii) utilizar vocabulario confuso o muy técnico, e (iii) imponer una carga mental excesiva a los usuarios. Para evitar estas situaciones, es necesario evaluar la usabilidad para que el usuario no se sienta insatisfecho con el producto final [Hertzum, 2020].

La realidad extendida (RX) es un concepto de hardware especializado que genera una experiencia con diferentes niveles de intensidad e interactividad dentro de entornos en los que los usuarios se sienten inmersos en un mundo aparentemente real y se define en función de la tecnología [Kardong-Edgren *et al.*, 2019; Tham *et al.*, 2018]. Actualmente, varios campos, como la educación [Hickman y Akdere, 2017], la salud [Heiyanthuduwa *et al.*, 2020] y la seguridad sísmica [Li *et al.*, 2017], utilizan la RX. Debido a la amplia gama de usuarios (niños, jóvenes, adultos, e incluso adultos mayores) que pueden interactuar con entornos de RX, es esencial considerar la usabilidad de este tipo de sistemas software, lo que aumenta la eficiencia y la satisfacción del usuario, mejorando también su productividad [Hertzum, 2020].

Los trabajos de investigación sobre la evaluación de la usabilidad de los entornos virtuales están dispersos en diferentes bases de datos científicas. Hasta donde

sabemos, ningún trabajo de investigación recopila e informa de forma global sobre qué técnicas se utilizan para evaluar la usabilidad de estos sistemas software. Por lo tanto, los profesionales e investigadores que necesiten esta información tendrán que: (i) Dedicar mucho tiempo a buscar y encontrar fuentes de conocimiento relevantes, (ii) revisar los recursos obtenidos según criterios específicos, (iii) extraer información de múltiples fuentes y, por último, (iv) concluir a partir de la información encontrada. Este problema se agudiza si se considera que la comunidad científica aporta periódicamente nuevos recursos bibliográficos.

El presente trabajo de título se desarrolla en el marco de una línea de investigación del profesor guía que pretende, entre otros aspectos, evaluar el panorama general de las técnicas relacionadas con la evaluación de la usabilidad que se están utilizando en entornos de RX. Para ello, realizamos un estudio de mapeo sistemático (EMS) abarcando desde 2018 hasta junio 2022. Las bases de datos utilizadas son Scopus, *IEEE Xplore* y *Web of Science* (WoS). Se identificaron un total de 30 estudios primarios que abordan la evaluación de entornos de RX. Entre estos estudios, se distinguen tres categorías principales de entornos: (1) Realidad Virtual (RV), que implica una inmersión total en un mundo ficticio; (2) Realidad Aumentada (RA), que mezcla objetos virtuales con el mundo real; y (3) Realidad Mixta (RM), que fusiona elementos virtuales y reales permitiendo interacciones más directas. De los 30 estudios primarios identificados, 25 se centran en la evaluación de sistemas de RV, mientras que sólo 3 abordan sistemas de RA y 2 se enfocan en sistemas de RM. Cada uno de estos tipos de entorno presenta particularidades únicas en su nivel de inmersión y relación con el entorno real.

Tras una revisión exhaustiva de los 30 estudios primarios se identificó que solo cuatro de ellos informan sobre problemas y retos al evaluar la usabilidad en entornos de RX. En dos de los estudios primarios [Granić *et al.*, 2020; Verkuyt *et al.*, 2018], se resaltan desafíos en relación con la muestra, como muestreo por conveniencia y altas tasas de abandono, junto con limitaciones de tiempo y recursos que restringieron rondas de evaluación. También se han identificado dificultades en la implementación de técnicas de evaluación en otros dos estudios primarios [Delamarre *et al.*, 2020; Kabassi *et al.*, 2018], incluyendo subjetividad en la evaluación heurística y ruptura de la sensación de presencia en entornos virtuales con la técnica de pensamiento en voz alta. Estos desafíos impactan la evaluación de usabilidad en entornos virtuales. A través del análisis de los estudios primarios, se identificaron

14 técnicas de evaluación de usabilidad presentes en la literatura del área de la Interacción Persona-Computador (IPC). Estas técnicas incluyen la entrevista, evaluación heurística, encuesta, experimento de usabilidad, revisión de expertos, recorrido cognitivo, retroalimentación del usuario, pensamiento en voz alta, grabaciones de pruebas de usabilidad, cuestionarios, Personas, seguimiento ocular, métricas de usabilidad y observación directa.

Luego del análisis de los estudios primarios se identificaron dos técnicas adaptadas para evaluar la usabilidad en entornos virtuales: encuestas y recorridos cognitivos. Las encuestas se utilizaron en diferentes contextos, abordando preguntas demográficas y aspectos de utilidad y facilidad de uso en una escala Likert. Por otro lado, la técnica de recorrido cognitivo adaptada implicó un análisis minucioso de tareas con evaluadores representando a diferentes usuarios, proporcionando información sobre objetivos, acciones y evaluación del sistema. Estas técnicas, a pesar de sus desafíos, brindaron valiosos conocimientos sobre la usabilidad en entornos virtuales. Además, en diversos estudios [Birnie *et al.*, 2018; Birt y Vasilevski, 2021; De Melo *et al.*, 2019; Doumanis y Economou, 2019; Kamińska *et al.*, 2022; Liu *et al.*, 2018; Ogunseiju *et al.*, 2022; Rohrbach *et al.*, 2019; Schmidt *et al.*, 2021; Tromp *et al.*, 2018; Verkuyl *et al.*, 2018], se han empleado combinaciones de técnicas para lograr una evaluación integral de la usabilidad en entornos virtuales. Se identificaron 12 combinaciones diferentes en el análisis de los estudios primarios, abarcando varias técnicas como entrevistas, evaluación heurística, encuestas y retroalimentación del usuario, entre otras. Estas combinaciones permiten obtener una visión más completa de la usabilidad en estos entornos.

También, en cinco estudios se han desarrollado *frameworks* para evaluar la usabilidad en entornos de RV. El primero se centra en la evaluación de la usabilidad en contextos artísticos de RV, considerando la respuesta emocional y el compromiso del espectador [He *et al.*, 2018]. Otro estudio utiliza un enfoque semiótico combinado con métodos cognitivos para evaluar la usabilidad en aplicaciones de RV [Barricelli *et al.*, 2018]. Un tercer *frameworks* [Liu *et al.*, 2018], se basa en las etapas de acción de Norman y busca mejorar la interfaz de RV mediante principios de diseño. Además, se presenta un *frameworks* de software llamado *Fishtank Sandbox* para evaluar técnicas de interacción en entornos de RV menos complejos [Jangid y Kong-silp, 2021]. Por último, se emplea un enfoque de métodos mixtos en la evaluación de un prototipo de *exergame* de RV, combinando técnicas cuantitativas y cualitativas

como el pensamiento en voz alta, cuestionarios y entrevistas [Thalman *et al.*, 2021]. Por último, dos estudios proponen nuevas técnicas para evaluar la usabilidad y la experiencia del usuario en entornos de RV. En el primer estudio [Granić *et al.*, 2020], se presenta un enfoque de evaluación de usabilidad en grupo basado en escenarios (ScerGUT, por sus siglas en inglés *Scenario-based Group Usability Testing*), que combina métodos como pruebas basadas en tareas, cuestionarios de actitud, pruebas de memorabilidad, entrevistas y observaciones del evaluador. El enfoque utiliza medidas cuantitativas y cualitativas para evaluar entornos virtuales de aprendizaje. En el segundo estudio [Harms, 2019], se propone un método automatizado que registra las acciones del usuario en la RV, genera árboles de tareas y detecta problemas de usabilidad a través de la identificación de “olores de usabilidad”. Este método se aplica en la RV mediante el registro de eventos específicos de interacción del usuario y análisis de árboles de tareas para detectar comportamientos problemáticos.

En este trabajo de título, se han identificado diversas técnicas comunes empleadas para evaluar la usabilidad en entornos de RX, tales como entrevistas, evaluación heurística y encuestas. Además, se han encontrado propuestas de *frameworks* destinados a lograr una evaluación más exhaustiva. Es importante resaltar que las entrevistas se destacan como una técnica especialmente valiosa para obtener perspectivas directas de los usuarios, mientras que la evaluación heurística se presenta como un enfoque sistemático para la mejora de las interfaces. Por otro lado, las encuestas, incluyendo el *System Usability Scale* (SUS), se utilizan para medir la satisfacción del usuario en estos contextos. La adaptación de técnicas, la combinación de diversos métodos y el desarrollo de *frameworks* específicos han sido observados en los estudios. Predominantemente, la mayoría de los estudios se han centrado en la evaluación de la usabilidad en entornos de RV, con una atención relativamente menor en los entornos de RA y RM. Se han identificado desafíos que abarcan problemas de muestreo y la aplicabilidad de técnicas en estos entornos. Además, se han discernido cinco enfoques distintos para la evaluación de la usabilidad en RX, entre los que se incluyen adaptaciones de técnicas, combinaciones de métodos, propuestas de *frameworks* y la implementación de métodos específicos. Estas prácticas se consideran fundamentales para la mejora constante de la experiencia del usuario en los entornos RX y para estimular la innovación en la industria. Como futuros trabajos, se propone la ampliación de la búsqueda y el análisis de la aplicación de técnicas

siguiendo los estándares de la IPC, junto con la consideración de las adaptaciones realizadas.

1.1. Objetivos

A continuación, se detalla el objetivo general y los objetivos específico planteados en esta investigación.

1.1.1. Objetivo General

Evaluar el panorama general de las técnicas relacionadas con la evaluación de la usabilidad que están siendo utilizadas en ambientes virtuales.

1.1.2. Objetivos Específicos

Los objetivos específicos del presente trabajo de título son los siguientes:

- a) Identificar las técnicas de usabilidad que están siendo utilizadas para evaluar ambientes virtuales,
- b) Establecer los problemas de usabilidad que presentan los ambientes virtuales,
- c) Determinar cómo se están aplicando las técnicas para evaluar la usabilidad de ambientes virtuales,
- d) Identificar los problemas y retos existentes en la aplicación de técnicas para evaluar la usabilidad de ambientes virtuales,
- e) Determinar los tipos de ambientes virtuales que se están evaluando.

1.1.3. Preguntas

Las preguntas que se plantean para el presente trabajo de título son:

1. ¿Cuáles son las técnicas de usabilidad que se están utilizando para evaluar entornos virtuales?
2. ¿Cuáles son los principales problemas y desafíos en la aplicación técnicas para evaluar la usabilidad de los entornos virtuales?

3. ¿Cuáles son los entornos de realidad extendida que se están evaluando actualmente?
4. ¿Cómo se aplican las técnicas de usabilidad para evaluar los entornos virtuales?

1.2. Metodología

La Ingeniería de Software considera la usabilidad como uno de los atributos clave del software [Castro *et al.*, 2022; Losana *et al.*, 2021]. La usabilidad es una medida que puede emplearse en cualquier tipo de producto, sin ser exclusivo de sistemas informáticos. El estudio secundario que se reporta en este trabajo se ha desarrollado siguiendo las directrices establecidas por [Kitchenham *et al.*, 2011] para la realización de un SMS. Siguiendo estas directrices, se llevaron a cabo las siguientes actividades:

- (1) Formular las preguntas: En esta etapa, se identifican las preguntas que guiarán el proceso de revisión. Estas preguntas deben ser claras y enfocadas, y aseguran que la búsqueda de estudios relevantes se centre en responder a preguntas específicas.
- (2) Definir la estrategia de búsqueda: Aquí se crea una estrategia exhaustiva para identificar estudios relevantes de diversas fuentes, como bases de datos. Esta estrategia garantiza la inclusión de todos los estudios relevantes y minimiza el riesgo de omitir información importante. Es importante desarrollar una estrategia efectiva que asegure la rigurosidad e imparcialidad del proceso de revisión.
- (3) Crear la cadena de búsqueda: En esta etapa se ajusta la cadena de búsqueda para asegurar su integralidad y precisión. Se realiza después de la búsqueda inicial y la revisión de los resultados en términos de relevancia y exhaustividad. La conformación de la cadena de búsqueda es un proceso iterativo que puede requerir varias rondas de refinamiento basadas en los comentarios de los revisores y búsquedas adicionales.
- (4) Definir los criterios de selección: Los criterios de inclusión y exclusión se eligen en función de las preguntas formuladas y deben ser definidos con claridad. Los criterios determinan los tipos de estudios que se incluirán o excluirán de la revisión. Es esencial establecer los criterios de manera anticipada para evitar sesgos y garantizar la transparencia en el proceso de selección. Además, se deben aplicar de manera consistente para asegurar la confiabilidad y validez de los resultados.

- (5) Seleccionar los estudios primarios: En esta etapa se evalúan los resultados de la búsqueda para identificar estudios relevantes que cumplan con los criterios de inclusión. Esta selección es fundamental para enfocarse en estudios de alta calidad que sean pertinentes para las preguntas u objetivos definidos previamente. El proceso puede requerir múltiples rondas de evaluación y es importante documentar todas las decisiones tomadas para garantizar transparencia y reproducibilidad.
- (6) Extracción y síntesis de la información: En esta etapa se extraen datos relevantes de los estudios seleccionados y se sintetizan para responder a las preguntas. Es crucial para asegurar que la revisión se base en evidencia de alta calidad y pertinente. El proceso puede requerir múltiples rondas de extracción y síntesis, y es importante documentar las decisiones tomadas para garantizar transparencia y reproducibilidad.

1.3. Estructura del Trabajo

El objetivo de esta sección es proporcionar una visión clara de la estructura y organización del trabajo de título, el cual se centra en el análisis del panorama actual de las técnicas utilizadas para evaluar la usabilidad de entornos de RX. En este sentido, se presenta una breve descripción de cada uno de los capítulos que conforman el presente trabajo, brindando una visión general de los contenidos y enfoques abordados en cada uno de ellos.

- 1) En el primer capítulo, que corresponde a la introducción, se presentan los objetivos, una visión clara de la metodología, estructura del trabajo de título y la publicación derivada.
- 2) En el segundo capítulo, se presenta el marco conceptual y teórico que sustenta el trabajo de titulación. Además, se lleva a cabo una revisión y análisis de estudios previos y trabajos relacionados que se centran en la evaluación de la usabilidad en entornos de RX. En este contexto, se destacan las diferencias y contribuciones específicas del trabajo de titulación.
- 3) En el tercer capítulo, se describe detalladamente la metodología utilizada, se explican los pasos y procedimientos empleados para recopilar y analizar los datos. Además, se realiza una síntesis de los hallazgos para proporcionar una visión general cohesiva y coherente de los resultados del estudio.

- 4) En el cuarto capítulo, se presentan y analizan detalladamente los resultados obtenidos a partir del estudio. En este proceso, se enfatizan los descubrimientos más significativos y se establecen conexiones claras con la literatura existente.
- 5) En el quinto capítulo, se discuten las implicaciones y posibles interpretaciones de los resultados. Además, se identifican y discuten las posibles limitaciones y amenazas para la validez del estudio.
- 6) En el sexto capítulo, se elaboran las conclusiones finales del trabajo, donde se da cumplimiento con los objetivos del trabajo de título, se resaltan las contribuciones significativas y se discuten las implicaciones clave de la investigación. Además, se identifican posibles direcciones futuras para la expansión y mejora de este estudio.
- 7) En el Anexo A, se presenta la publicación derivada del presente trabajo de titulación y en el Anexo B, se muestran los estudios primarios identificados a través de la metodología seguida.

1.4. Publicación Derivada

En el contexto del trabajo de titulación, se llevó a cabo un estudio exploratorio acerca de las técnicas de evaluación de usabilidad para entornos virtuales, la cual dio como resultado la elaboración de un artículo titulado *Usability Evaluation Techniques for Virtual Environments: An Exploratory Study*. Este artículo fue aceptado y presentado en la *25th International Conference on Human-Computer Interaction (HCI International 2023)*, que se llevó a cabo en el AC Bella Sky Hotel y Bella Center, en Copenhague, Dinamarca, aunque fue presentado de manera remota. La publicación derivada se presenta en el Anexo A.

Capítulo 2

Marco Teórico y Trabajos Relacionados

En este capítulo, se abordará el marco teórico y los trabajos relacionados. En el marco teórico se sustenta el trabajo de titulación, con el propósito de sentar las bases conceptuales que permitan analizar el problema planteado y dar cumplimiento a los objetivos propuestos. Para ello, se hará una revisión de los conceptos relevantes en el campo de estudio. Por otro lado, los trabajos relacionados expondrán los antecedentes vinculados al tema central de este trabajo de título.

2.1. Usabilidad

El término “usabilidad” surgió en 1971 y ha sido desarrollado por diversos autores a lo largo del tiempo para describir este atributo de calidad. Sin embargo, estas definiciones presentaban una limitación, ya que no proporcionaban una precisión cuantificable de lo que implica la usabilidad [Shackel, 2009]. En la actualidad, la usabilidad se define según la norma ISO 9241 (parte 11) como la capacidad de un producto para ser utilizado por usuarios específicos con el propósito de alcanzar objetivos de manera efectiva, eficiente y satisfactoria en un contexto determinado, con el respaldo de usuarios específicos y para cumplir las tareas para las que fue diseñado [Ferré *et al.*, 2001; Hertzum, 2020; Shackel, 2009].

Es importante destacar que la usabilidad no se limita únicamente a la interfaz de usuario, sino que también abarca la estructura general del software y el concepto subyacente. Lograr la usabilidad en un sistema plantea un desafío, ya que requiere

conocimientos especializados y una profunda comprensión de las preferencias, requisitos y limitaciones de los usuarios. Es fundamental considerar la usabilidad como un atributo de calidad durante el desarrollo de software, aunque también se deben tener en cuenta otros atributos. No obstante, integrar la usabilidad en el proceso de desarrollo no es una tarea sencilla y puede generar conflictos entre los equipos de usabilidad y desarrollo debido a sus diferentes perspectivas. Tradicionalmente, la usabilidad se abordaba tardíamente en el ciclo de desarrollo, lo que resultaba en costos elevados para corregir problemas significativos [Juristo *et al.*, 2001].

2.2. Evaluación de la Usabilidad

La evaluación de la usabilidad desempeña un papel central en el proceso de usabilidad, ya que permite determinar el nivel de usabilidad de la versión actual del sistema software y verificar si el diseño cumple su función. Entre los diversos métodos utilizados para llevar a cabo esta evaluación, las pruebas de usabilidad son especialmente relevantes. Estas pruebas implican la participación de un grupo de usuarios que interactúan con el sistema, mientras se registran y analizan meticulosamente los resultados obtenidos. Es esencial destacar que no es posible predecir la usabilidad de un sistema de software sin someterlo a pruebas con usuarios reales, ya que sus opiniones y experiencias son fundamentales en este proceso de evaluación [Ferré *et al.*, 2001].

Mediante la evaluación de la usabilidad, se evitan situaciones indeseables y se mejora la experiencia del usuario al identificar y resolver problemas durante la interacción con el producto [Hertzum, 2020]. Al realizar una prueba de usabilidad, es importante evitar ciertos errores comunes que pueden afectar su efectividad y la información obtenida sobre la experiencia del usuario. Algunos de estos errores incluyen: probar con usuarios equivocados o en un número insuficiente, proporcionar demasiada orientación a los participantes, no establecer metas y objetivos claros para la prueba, no evaluar las tareas o escenarios adecuados, y no registrar y analizar los datos de manera efectiva. Al evitar estos errores, los evaluadores de usabilidad pueden asegurarse de que sus pruebas sean más efectivas y brinden información valiosa para mejorar la experiencia del usuario [Hertzum, 2020].

2.3. Tipos de Entornos

Para hablar sobre los distintos tipos de entornos, es importante comprender el término de RX. Este término hace referencia a una variedad de tecnologías y fenómenos que implican la interacción entre espacios digitales y materiales. En otras palabras, la RX engloba la RV, la RA y la RM [Rzeszewski y Orylski, 2021].

2.3.1. Realidad Virtual

La RV se define como un hardware tecnológico que crea una experiencia inmersiva en un entorno virtual, brindando diferentes niveles de interactividad. Se basa principalmente en la visión y la manipulación manual para generar una sensación de presencia en el espacio virtual. Para su implementación, se utilizan gafas de visualización estéreo y guantes de realidad. Además, la RV se caracteriza por utilizar características 3D inmersivas y altamente visuales para reproducir situaciones de la vida real. A diferencia de la simulación basada en computadora, la RV incorpora interfaces físicas o digitales, como teclados, ratones, reconocimiento de voz, sensores de movimiento y dispositivos hápticos.

En términos generales, el entorno de RV se define como una realidad ficticia en la que el usuario está completamente inmerso y puede interactuar. Este mundo virtual puede simular propiedades de entornos reales existentes o ficticios. Para sumergirse en la RV, se requieren dispositivos de visualización, conocidos como pantallas montadas en la cabeza (PMC), los cuales también funcionan como barreras para mantener al usuario dentro de este entorno virtual [Hickman y Akdere, 2017; Kardong-Edgren *et al.*, 2019; Tham *et al.*, 2018]. La Figura 2.1 muestra como es un entorno de RV.



(a) HTC Vive

(b) Usuario usando un HTC Vive

Figura 2.1: Entornos de realidad virtual, (a) HTC Vive, (b) Usuario usando un HTC Vive. Tomado de Tham *et al.* (2018)

2.3.2. Realidad Aumentada

La RA se define como una variante de la RV en la que se superponen estímulos sintéticos a objetos del mundo real, con el propósito de hacer perceptible información que de otra manera sería imperceptible para los sentidos humanos. Es una tecnología que combina información digital generada por computadora con objetos y lugares reales, con el objetivo de mejorar la experiencia del usuario. La RA tiene como objetivo principal mejorar el proceso de aprendizaje y se sitúa en un punto intermedio entre el mundo real y el mundo virtual, formando parte del espectro de simulación de la RM. Al combinar elementos virtuales con el entorno físico, permite una interacción más inmersiva y enriquecedora. Esto se logra mediante el uso de dispositivos como gafas de RA, que superponen información digital en tiempo real sobre el entorno visual del usuario, brindando así una experiencia única y enriquecedora [Azuma *et al.*, 2001; Kardong-Edgren *et al.*, 2019]. La Figura 2.2 muestra como es un entorno de RA.

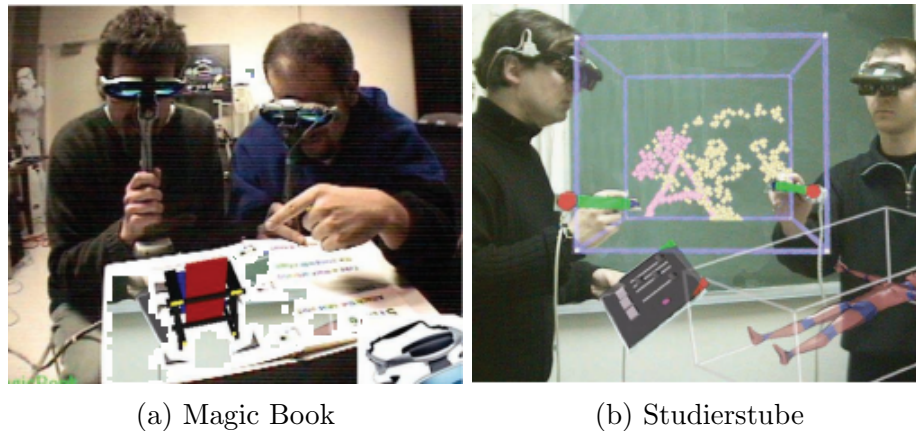


Figura 2.2: Entornos de realidad aumentada, (a) Magic Book, (b) Studierstube, son sistemas de RA colaborativos, con dos usuarios que llevan PMC transparentes según Azuma *et al.* (2001)

2.3.3. Realidad Mixta

La RM se sitúa en un punto intermedio del espectro de la simulación de la realidad, entre el mundo real y el mundo virtual. Es una fusión de la RA y la RV, donde los estímulos sintéticos se superponen a objetos del entorno físico. A diferencia de la RA, que combina elementos reales y virtuales en tiempo real, la RM

va más allá al establecer una integración fluida entre ellos. Esta tecnología no solo superpone elementos virtuales al entorno físico, sino que también crea una relación y alineación precisa entre los objetos reales y virtuales. Mediante el uso de dispositivos como las gafas de RM, los usuarios experimentan una interacción más inmersiva y enriquecedora, donde los objetos virtuales pueden interactuar con el entorno real y viceversa. Además, en la RM, los objetos digitales están vinculados al espacio real de manera precisa. Por ejemplo, un holograma puede ser superpuesto a un objeto concreto, mientras que en la RA, un holograma de un objeto específico puede colocarse en cualquier superficie plana [Kardong-Edgren *et al.*, 2019]. La Figura 2.3 muestra como es un entorno de RM.



(a) Ritchie

(b) Humanoide Virtual

Figura 2.3: Entornos de realidad mixta, (a) Ritchie, (b) Humanoide Virtual, son agentes de realidad mixta reportados por Holz *et al.* (2011)

2.4. Trabajos Relacionados

Como resultado de la búsqueda piloto realizada al inicio del trabajo de titulación, encontramos tres estudios de revisión sistemática de literatura [Karre *et al.*, 2020; Sheikh *et al.*, 2021; Zanatta *et al.*, 2022] relacionados con el objetivo. El trabajo de [Karre *et al.*, 2020] reporta estudios de usabilidad realizados en la industria cuando se desarrollan productos de RV. Los autores se centran en aquellos estudios que diseñan escenas de RV de forma precisa, y elaboran un protocolo sencillo para realizar o establecer una configuración para la evaluación de la usabilidad de dichas escenas. Este estudio pretende servir de contribución al momento de escoger un método de

evaluación de la usabilidad para futuros productos de RV a nivel industrial. A pesar de que este estudio es completo, se encuentra desactualizado, ya que, se realiza desde el 2000 hasta el 2018.

En el segundo trabajo [Sheikh *et al.*, 2021], se reporta del uso de ciertas técnicas de evaluación de usabilidad en sistemas de RA específicamente en el área de la educación. Los autores aseguran que para implementar la RA es necesario tener presente algunos retos importantes, especialmente en el área educativa, donde se puede requerir de una mayor demanda de habilidades motoras por partes de los usuarios debido a la interacción con proyecciones en 3D. En este trabajo, los autores reportan una miscelánea de técnicas de usabilidad que son eficientes a la hora de evaluar sistemas de RA para el área de la educación. Aunque esto claramente podría extenderse a otras áreas que hagan uso de tal tecnología, no se centra en todas las áreas aplicables de la tecnología de RA.

En el último estudio, [Zanatta *et al.*, 2022] reportan trabajos donde se evalúa la usabilidad de sistemas de RV y dispositivos robóticos aplicados a la rehabilitación neuromotora. Los autores se enfocan en una variedad de ámbitos importantes para poder implementar un sistema de RV en el área de la rehabilitación, entre los cuales se menciona la usabilidad, y se abordan las técnicas de usabilidad que fueron visualizadas en los estudios primarios identificados en su investigación. Aunque el estudio de [Zanatta *et al.*, 2022] entrega información actualizada y completa, y entrega algunas recomendaciones, no es suficientemente genérico, ya que, este se centra en un área de las tantas a las que se aplica la RV.

Luego de realizar un análisis de los estudios mencionados anteriormente, encontramos que el presente trabajo de titulación difiere de los dos últimos (i.e., [Sheikh *et al.*, 2021] y [Zanatta *et al.*, 2022]), ya que, no nos limitaremos a áreas particulares, y consideraremos no solamente la RV sino también RA y RM con el objetivo de obtener una visión más amplia de cómo se está evaluando la usabilidad de los sistemas que usan estas tecnologías. Asimismo, nuestro trabajo difiere del primer estudio descrito (i.e., [Karre *et al.*, 2020], puesto que, realizamos una investigación actualizada (desde el 2018 hasta junio del 2022), y el enfoque de las preguntas es diferente. Por tanto, el presente trabajo de investigación es necesario ya que el uso de las tecnologías RV, RA y RM actualmente está en constante creciente y se requiere conocer cuáles son las técnicas utilizadas para evaluar su usabilidad de manera correcta.

Capítulo 3

Metodología y Síntesis

En este capítulo, se aborda la metodología utilizada en el estudio secundario realizado en este trabajo de título y se ofrece una síntesis concisa de los resultados obtenidos. La metodología desempeña un papel crucial al proporcionar los medios para obtener datos pertinentes y confiables que contribuyen al logro de los objetivos planteados. El estudio secundario que se presenta en este trabajo ha seguido rigurosamente las directrices establecidas por [Kitchenham *et al.*, 2011] para llevar a cabo un EMS. Estas directrices comprenden varias etapas críticas, como la formulación de las preguntas de investigación, la definición de la estrategia de búsqueda, la selección de los estudios primarios, la extracción de datos y la síntesis de estos. A través de esta síntesis, se busca proporcionar una visión panorámica de los hallazgos clave y destacar su relevancia en el contexto de la usabilidad en entornos de RX, sentando así las bases necesarias para un análisis en profundidad de los resultados obtenidos.

3.1. Identificación de los Estudios Primarios

El presente trabajo se basa en la metodología EMS, tal como se describe en la Sección 1.2. A continuación, se presentan los resultados obtenidos al aplicar dicha metodología, siguiendo los pasos desde la definición de la estrategia de búsqueda hasta la extracción de información y síntesis de resultados, tal como se ha propuesto en [Kitchenham *et al.*, 2011].

3.1.1. Definición de la Estrategia de Búsqueda

El EMS inicia con la identificación de las palabras clave, por lo que es necesario identificar un grupo de control (GC). El GC se define como un conjunto de estudios relacionados específicamente con el área de interés. Además, los estudios pertenecientes al GC deben representar con la mayor precisión posible la investigación, así como, responder las preguntas formuladas [Zhang *et al.*, 2011]. El GC nace de la necesidad de objetividad en la selección de la cadena de búsqueda.

Para formar el GC, se realiza una búsqueda manual de estudios relacionados con el objetivo del trabajo de titulación, asimismo, deben responder las preguntas planteadas. Como resultado de la búsqueda, se identificaron cinco estudios [Besoain *et al.*, 2018; Naranjo *et al.*, 2019; Oliveira *et al.*, 2017; Özding *et al.*, 2016; Pedrolí *et al.*, 2018]. Antes de construir la cadena de búsqueda, se verifica que los estudios del GC se encuentran en la base de datos de Scopus, ya que, es la base de datos con mayor cantidad de estudios. Como resultado de esta verificación, se encuentran todos los estudios del GC en Scopus.

Para obtener las palabras clave que conformarán la cadena de búsqueda se utiliza el programa Atlas.ti (versión 22), a partir del cual se generó un tabla con la frecuencia de todas las palabras y combinaciones de palabras que aparecían en los artículos del GC. Se escogieron solo aquellas palabras que se encontraban directamente relacionadas con las preguntas planteadas y que estuvieran presentes en un porcentaje significativo de los artículos pertenecientes al GC (a este porcentaje le denominamos cobertura).

Posteriormente, a cada una de las palabras obtenidas se le asignó un peso de 0 a 1, determinado por su frecuencia de uso, de forma que la palabra con mayor frecuencia de uso en los estudios del GC tuviera el peso 1. Se consideraron solamente las palabras con un peso igual o mayor a 0.327. La Tabla 3.1 muestra la lista de palabras obtenida como resultado de este proceso de selección. En ella, se muestran las palabras, su cobertura, frecuencia de uso y peso asignado (calculado con base en la cobertura y la frecuencia de uso – ver Ecuación 3.1). La síntesis de los pasos previamente explicados se muestran en la Figura 3.1.

$$Peso = \frac{\frac{Cobertura\ de\ la\ Palabra}{Maxima\ Cobertura} + \frac{Frecuencia\ de\ la\ palabra}{Maxima\ Frecuencia}}{2} \quad (3.1)$$

Tabla 3.1: Tabla de frecuencia de palabras obtenidas

Palabras	Cobertura (%)	Frecuencia	Peso
Usability	100	222	1
Evaluation	100	123	0.777
VR	100	101	0.727
Virtual Reality	100	90	0.703
Study	100	59	0.633
Virtual Environment	100	49	0.610
Test	100	34	0.577
Studies	80	28	0.463
Evaluations	60	38	0.386
Assesment	60	19	0.343
Immersive	60	18	0.341
Tests	60	12	0.327

3.1.2. Creación de la Cadena de Búsqueda

Luego de haber obtenido las palabras clave, se construyeron varias cadenas de búsqueda para lo cual se consideraron tres componentes relacionados con el objetivo del trabajo (es decir, identificar el panorama actual de cómo se realiza la evaluación de usabilidad de los entornos virtuales). Los componentes definidos fueron los siguientes: (i) evaluación, (ii) usabilidad y (iii) entornos virtuales. Se utilizó el operador lógico AND para unir cada uno de los componentes y OR para incluir sinónimos de los términos de un mismo componente.

En total, se construyeron seis cadenas para buscar los estudios del GC en la base de datos de Scopus y determinar cual de las cadenas es la mejor (es decir, aquella que encuentra todos los estudios del GC en el universo más pequeño de resultados). La Tabla 3.2 muestra el número de estudios que fueron encontrados y el número de estudios del GC encontrados por cada cadena probada.

Como se observa en la Tabla 3.2, cinco de las seis cadenas de búsqueda encontraron los cinco estudios del GC, haciendo necesario buscar un método para seleccionar una de las cadenas. Este método consiste en el cálculo de las proporciones X (véase la Ecuación 3.2) y Y (véase la Ecuación 3.3), y la media entre ambas proporciones (véase la Ecuación 3.4).

$$Xratio = \frac{\text{No de articulos encontrados en el grupo de control}}{\text{Total de articulos en el grupo de control}} \quad (3.2)$$

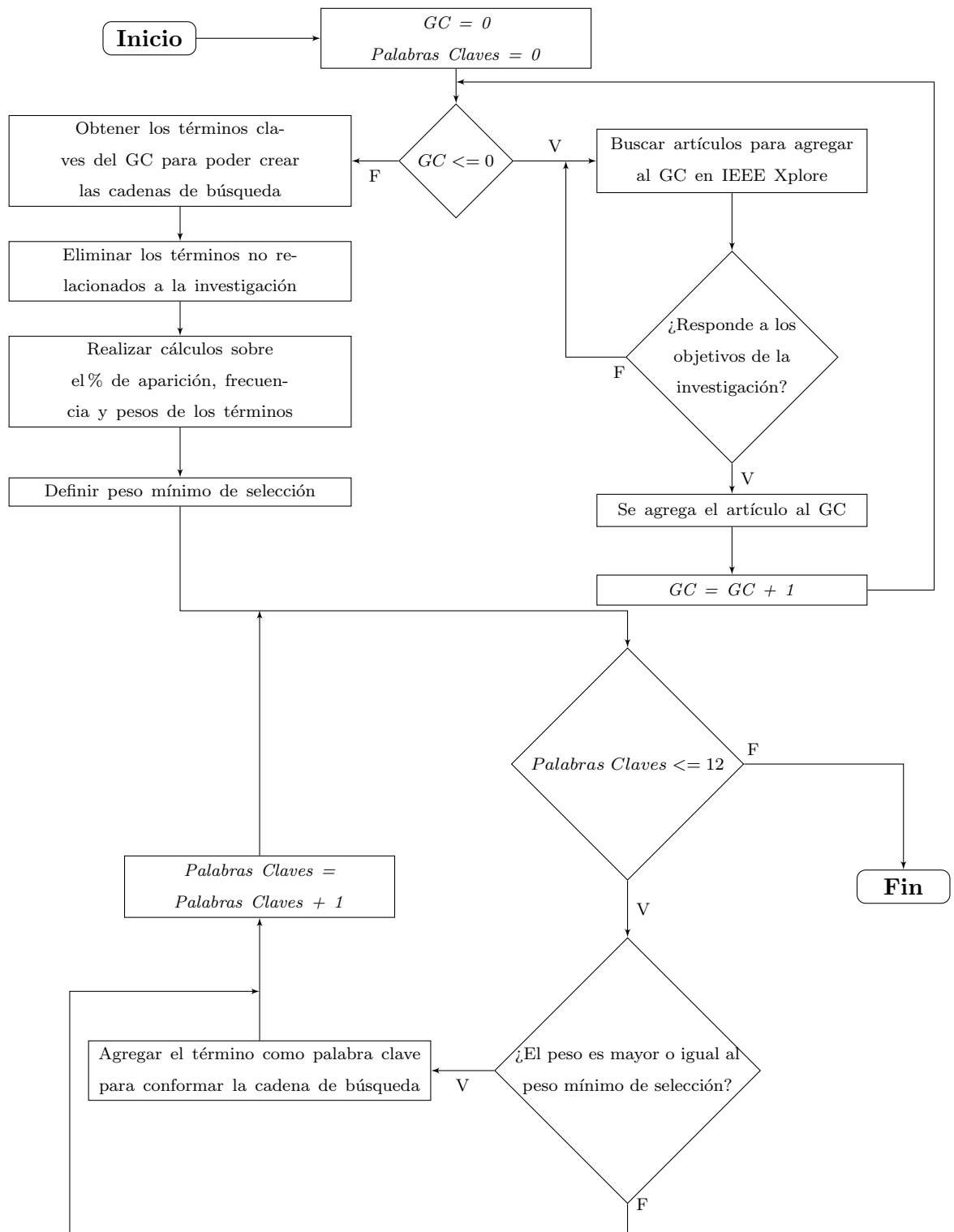


Figura 3.1: Tareas realizadas para obtener las palabras claves

Tabla 3.2: Cadena de palabras

ID	Cadena de búsqueda	Estudios encontrados	GC encontrados	Ratio X	Ratio Y	Promedio
1	(study OR evaluation OR studies OR assessment OR evaluations OR test) AND usability AND (VR OR “virtual reality” OR “virtual environment” OR immersive)	1955	5	1	0.00256	0.50128
2	(study OR evaluation OR assessment OR test) AND usability AND (VR OR “virtual reality” OR “virtual environment” OR immersive)	1801	5	1	0.00278	0.50139
3	(study OR evaluation OR assessment OR test) AND usability AND (VR OR “virtual reality” OR “virtual environment”)	1833	5	1	0.00273	0.50136
4	(study OR evaluation OR assessment) AND (usability) AND (VR OR “virtual reality” OR immersive)	1777	5	1	0.00281	0.50141
5	(study OR evaluation OR assessment) AND (usability) AND (VR OR “virtual reality”)	1666	4	0.8	0.00300	0.40150
6	(study OR evaluation OR studies OR assessment OR evaluations OR test)AND usability AND (VR OR “virtual reality”)	1764	5	1	0.00284	0.50142

$$Yratio = \frac{\text{No de articulos encontrados en el grupo de control}}{\text{Total de articulos encontrados por la cadena de busqueda}} \quad (3.3)$$

$$Promedio = \frac{X Ratio + Y Ratio}{2} \quad (3.4)$$

Como se observa en la Tabla 3.2, la proporción X varía solo una vez, ya que solo una de las cadenas no encontró todos los estudios del GC en Scopus. Sin embargo,

la proporción Y muestra diferencias específicas, ya que se basa en el cálculo de la proporción de los estudios del GC encontrados en el total de los resultados obtenidos por cada cadena. Para asegurarnos que la cadena seleccionada es la ideal para nuestro trabajo, se calcula la media entre la proporción X y la proporción Y. Como muestra la Tabla 3.2, la cadena 6 tiene la media más alta, por lo que se selecciona como la cadena más adecuada. A continuación, podemos visualizar la cadena final: *(study OR evaluation OR studies OR assessment OR evaluations OR test) AND (usability) AND (VR OR “virtual reality”)*

Una vez definida la cadena, se realizaron búsquedas en las bases de datos *IEEE Xplore* y *WoS* para complementar los resultados obtenidos inicialmente en *Scopus*. En esta búsqueda, se consideraron estudios desde el 2018 hasta junio del 2022. Las bases de datos se analizaron secuencialmente, utilizando los campos de búsqueda que se muestran en la Tabla 3.3. Los campos de búsqueda utilizados estuvieron determinados por las opciones proporcionadas por cada base de datos, debido a las diferentes sintaxis de consulta [Castro *et al.*, 2022; Magües *et al.*, 2016; Ren *et al.*, 2019]. Si aparecía un duplicado, se conservaba el primer resultado. En la Figura 3.2, se resume el proceso seguido para obtener los terminos claves que conforman la cadena de búsqueda.

Tabla 3.3: Campos de búsqueda por base de datos

Base de datos	Campos de búsqueda	Número de resultado
Scopus	“Title OR Abstract OR Keywords”	1425
IEEE Xplore	“Abstract”	198
Web of Science	“Title OR Abstract OR Keywords”	1035

3.1.3. Definición de los Criterios de Selección

En esta sección, se presentan los criterios de inclusión y exclusión para realizar la selección de los estudios primarios. Los criterios de inclusión son los siguientes:

1. El estudio indica las técnicas utilizadas durante la evaluación de usabilidad; AND
2. El estudio reporta con detalle la evaluación de usabilidad o un estudio de usabilidad; OR
3. El estudio reporta de manera detallada una propuesta de método/ técnica/ herramienta para evaluar la usabilidad de sistemas de RX.

Con respecto a los criterios de exclusión, son los siguientes:

1. El estudio solo utiliza una encuesta o cuestionario para realizar la evaluación de usabilidad.
2. El estudio no menciona la técnica de evaluación de usabilidad utilizada en el resumen.
3. El estudio reporta pruebas de usabilidad piloto o preliminares.

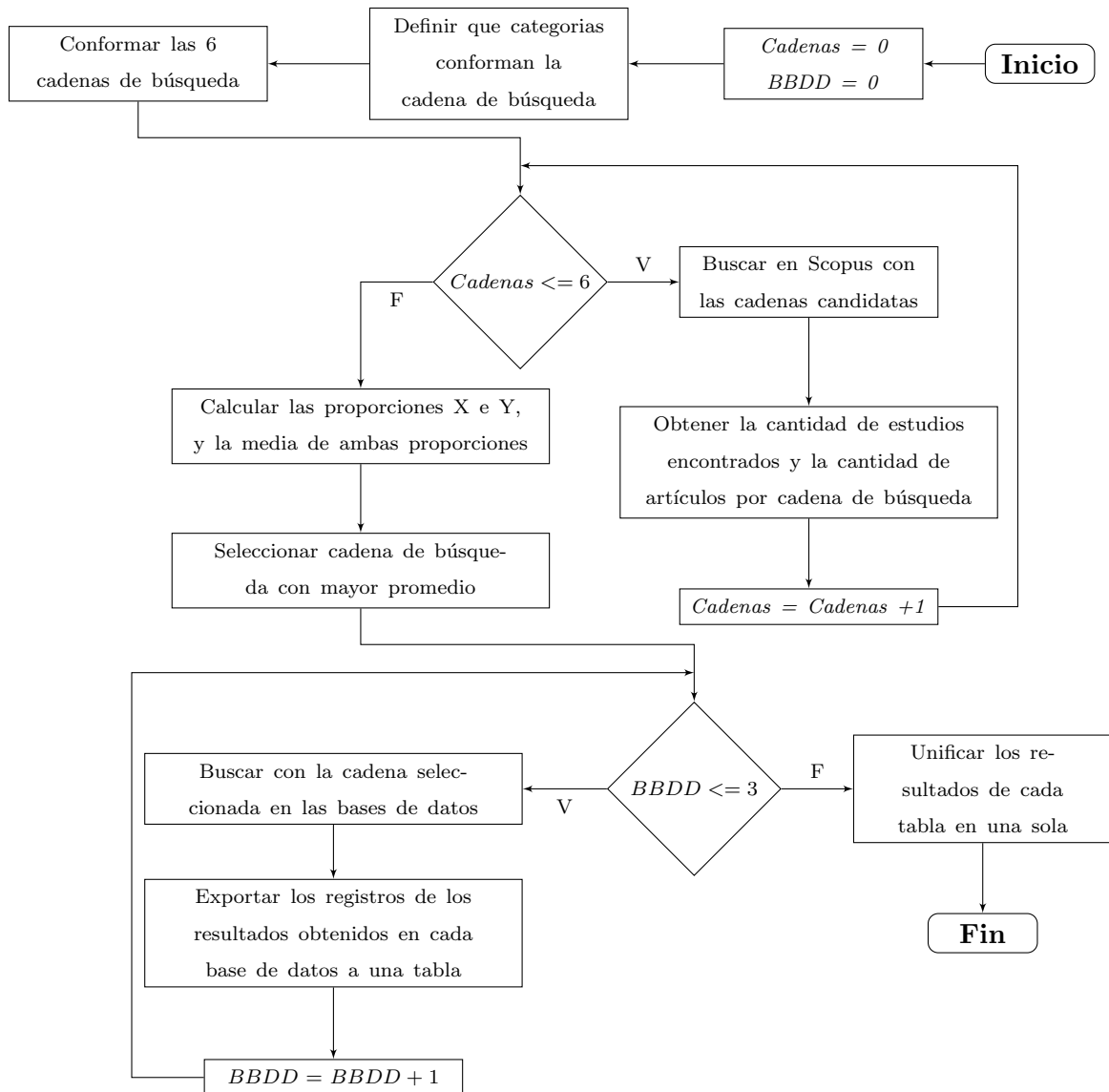


Figura 3.2: Tareas realizadas para conseguir los estudios preseleccionados de las diferentes bases de datos

4. El estudio está escrito en un idioma diferente al inglés.

Se debe considerar que es suficiente con que un estudio cumpla con uno de los criterios de exclusión para no ser considerado.

3.1.4. Selección de los Estudios Primarios

Se encontraron un total de 2658 estudios en las tres bases de datos utilizadas. Tras excluir los estudios duplicados, se redujo el número a 1781. Luego, se realiza una selección de estudios aplicando los criterios de inclusión y exclusión al título y resumen de cada uno de los estudios no duplicados, reduciendo el número a 289 preseleccionados. Finalmente, se realizó una aplicación estricta de los criterios de selección al texto completo de los estudios preseleccionados. La Figura 3.3 muestra todo el proceso de filtrado y análisis con los criterios de inclusión y exclusión utilizados para seleccionar finalmente 30 estudios primarios. La lista completa de los estudios primarios se encuentra en el Anexo B. Los resultados de la aplicación de los diferentes filtros durante el proceso de selección para cada base de datos pueden verse en la Tabla 3.4.

Tabla 3.4: Número de estudios restantes después de filtrar los resultados de la base de datos

Base de datos	Estudios encontrados	Estudios sin duplicados	Estudios preseleccionados	Estudios primarios
Scopus	1425	1418	244	25
IEEE Xplore	198	26	8	0
Web of Science	1035	337	37	5
Total	2658	1781	289	30

3.1.5. Extraer y Sintetizar la Información

En la etapa final, se procede a descargar los estudios primarios con el fin de crear una tabla Excel en la cual se recopila la siguiente información: título del estudio, autores, tipo de entorno de RX evaluado, técnicas utilizadas para evaluar la usabilidad, resultados obtenidos al aplicar dichas técnicas, adaptaciones o combinaciones de las técnicas, y *frameworks* reportados.

Posteriormente, se lleva a cabo la lectura y análisis de los 30 estudios primarios con el propósito de extraer la información necesaria para completar la tabla Excel

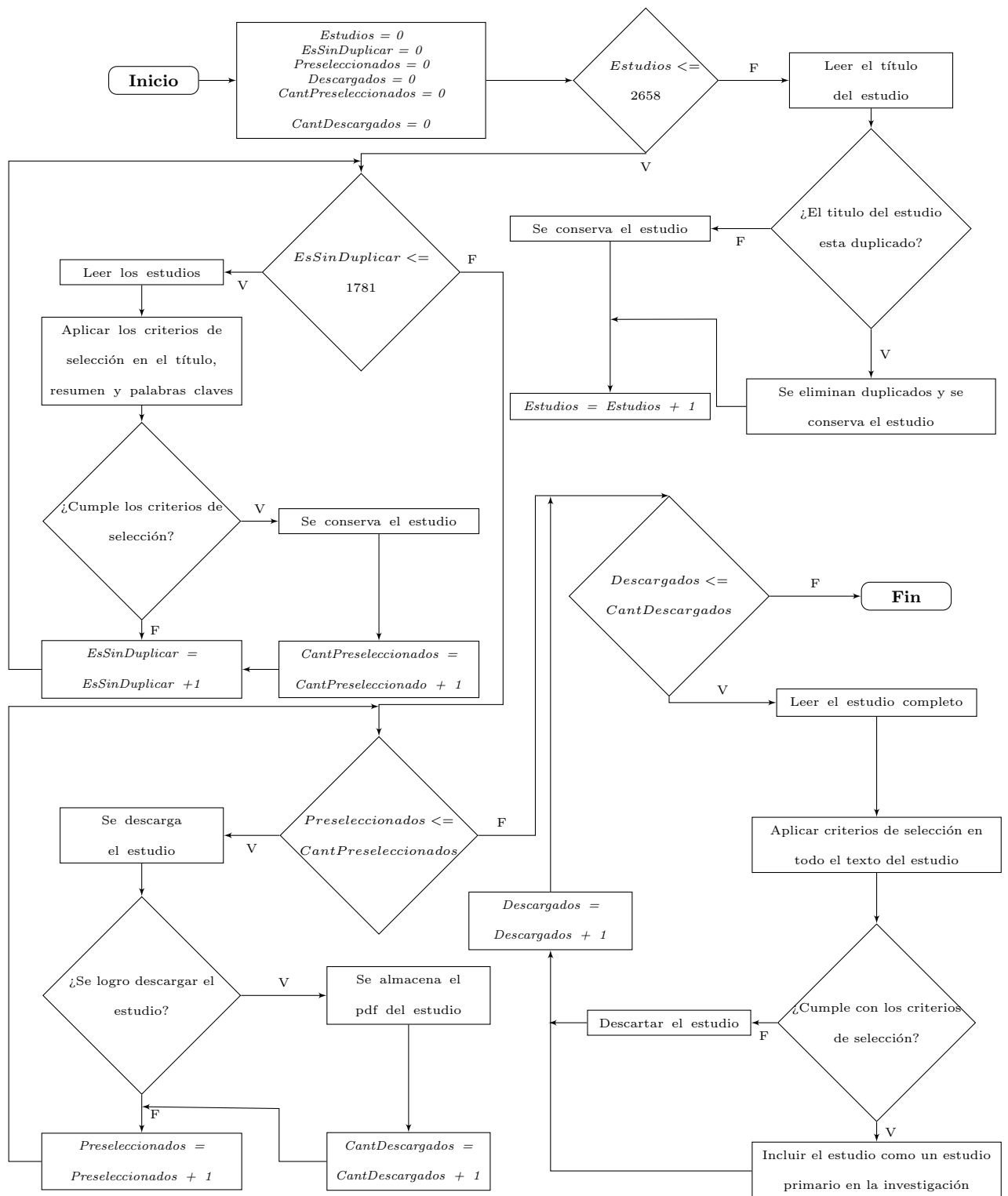


Figura 3.3: Tareas realizadas en el proceso de selección de los estudios primarios

mencionada anteriormente. A continuación, se realiza una síntesis de la información relevante con el fin de responder a las preguntas planteadas en la Sección 1.1.3. En la Figura 3.4 se resume visualmente el trabajo realizado en esta sección.

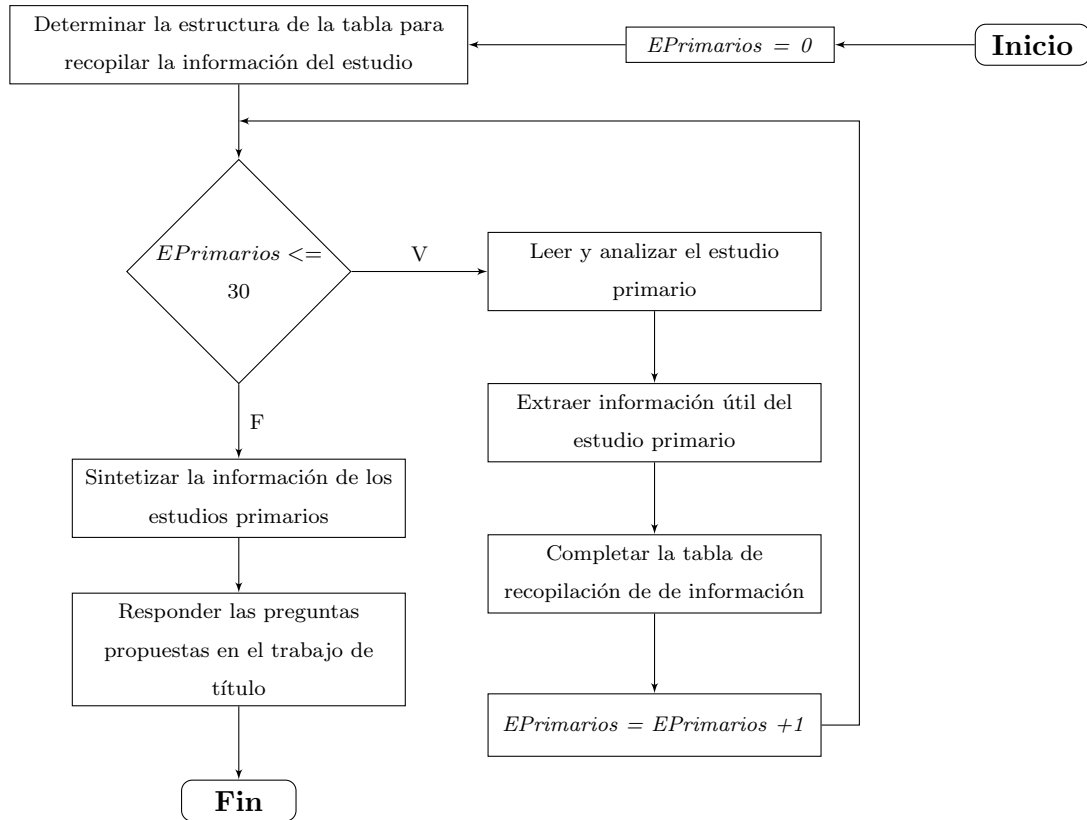


Figura 3.4: Tareas realizadas para extraer y sintetizar la información de los estudios primarios

3.2. Síntesis de los Estudios Primarios

En la Figura 3.5, se resume de manera concisa la información obtenida de los estudios primarios identificados. La Figura 3.5 se divide en dos áreas principales que ofrecen una visión global de los resultados. En el lado izquierdo, se encuentran dos gráficos de dispersión XY con burbujas, las cuales representan las intersecciones de las categorías de tipo-año de publicación (parte superior) y el tipo de publicación-técnicas/*framework* utilizados para evaluar la usabilidad (parte inferior). Los tipos de publicación incluyen revistas, capítulos de libro, conferencias y *workshops*, y el tamaño de cada burbuja refleja el número de estudios primarios clasificados en cada

categoría. En el lado derecho de la Figura 3.5, se muestra el número de estudios primarios según el año de publicación.

En la parte superior izquierda, se observa que las revistas son la categoría predominante, seguidas por las conferencias, capítulos de libro y *workshops*. Es destacable el aumento significativo que experimentaron las publicaciones en revistas en el año 2021. En la parte inferior izquierda, se identifican las técnicas de usabilidad utilizadas, entre las cuales podemos observar aquellas que han sido más empleadas, tales como encuestas, entrevistas, evaluación heurística y *framework* propuestos por los autores de los estudios primarios. En la parte superior derecha, se muestra la tendencia en la evaluación de la usabilidad de los entornos virtuales a lo largo del tiempo. Se aprecia una disminución en el interés durante los años 2019 y 2020, seguida de un aumento en 2021 y una ligera disminución en 2022 (esta disminución se debe a que para el 2022 se consideraron las publicaciones hasta junio).

En las siguientes cuatro secciones, se responde las preguntas planteadas en el trabajo de titulación, a partir del análisis de los estudios primarios identificados en el EMS.

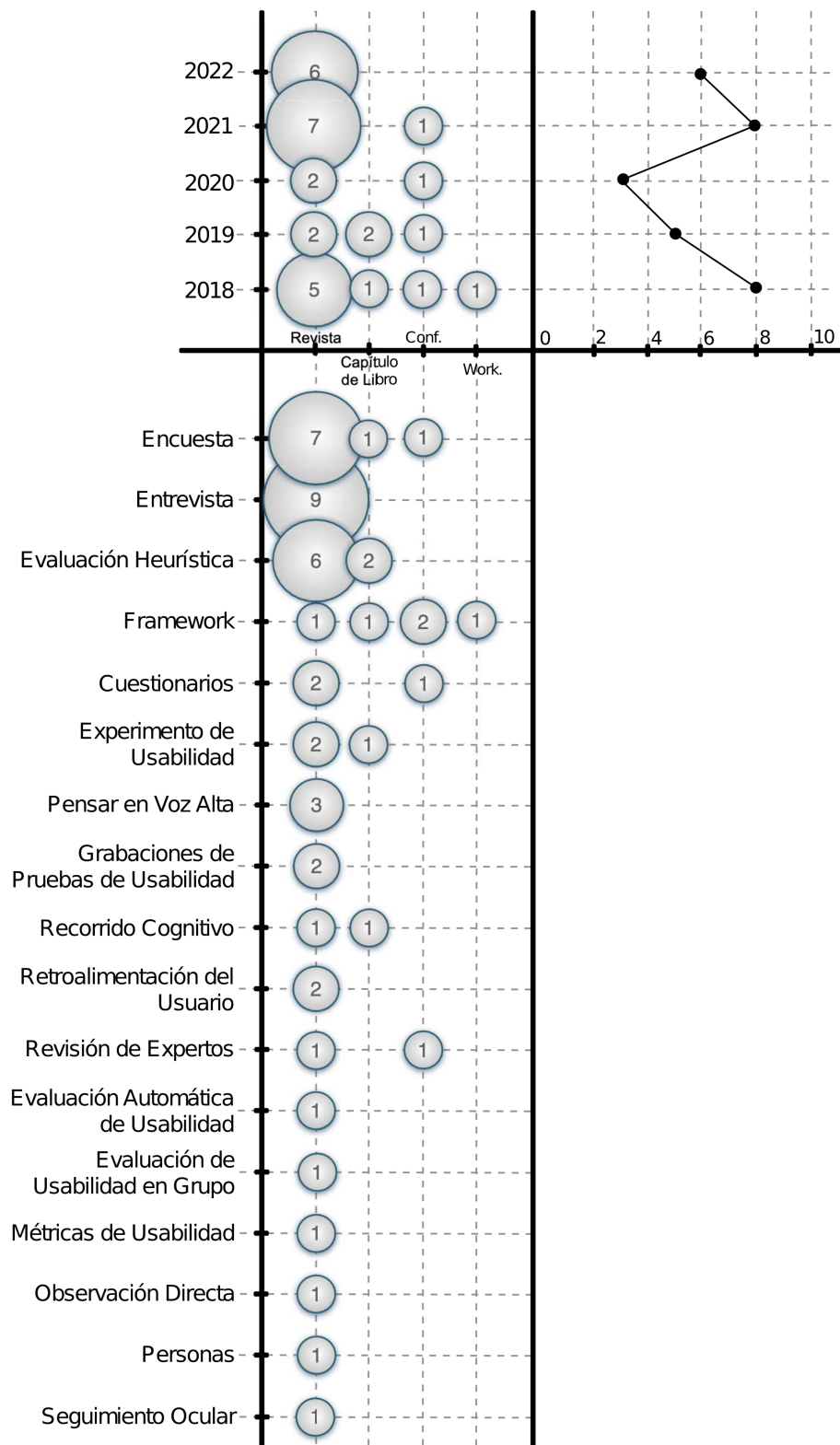


Figura 3.5: Etapas de la metodología a utilizar (EMS)

Capítulo 4

Análisis de Resultados

En este capítulo, se presentará el análisis de resultados tras procesar los datos recolectados, con el propósito de responder a las preguntas planteadas. Se detallarán minuciosamente los hallazgos más significativos que surgieron durante la fase de análisis de datos, con el fin de ofrecer una información clara y precisa acerca de la interpretación de los resultados y su relación con los objetivos planteados.

4.1. Técnicas de Usabilidad para Evaluar Entornos Virtuales

En esta sección, se da respuesta a la primera pregunta de investigación *¿Cuáles son las técnicas de usabilidad que se están utilizando para evaluar entornos virtuales?* Del análisis de los estudios primarios, se identificaron 14 técnicas de evaluación de usabilidad reportadas por la literatura de IPC: entrevista, evaluación heurística, encuesta, experimento de usabilidad, revisión de expertos, recorrido cognitivo, retroalimentación del usuario, pensamiento en voz alta, grabaciones de pruebas de usabilidad, cuestionarios, Personas, seguimiento ocular, métricas de usabilidad y observación directa. A continuación, daremos una breve descripción de las técnicas reportadas.

La **entrevista** es una técnica ampliamente utilizada para la recolección de información entregada por el usuario o los evaluadores de la usabilidad de los sistemas de RX [Nielsen, 1994b; Preece *et al.*, 1994]. Esta técnica ha sido utilizada y reportada en 9 de los estudios primarios [Birnie *et al.*, 2018; Cheng, 2022; Hassandra *et al.*,

2021; Kamińska *et al.*, 2022; Pedroli *et al.*, 2018; Rohrbach *et al.*, 2019; Rzeszewski y Orylski, 2021; Schmidt *et al.*, 2021; Verkuyl *et al.*, 2018]. Las entrevistas se utilizan para evaluar diversos tipos de entornos de RX. Por un lado, se han identificado 7 estudios primarios que se centran en la evaluación de la usabilidad en entornos de RV [Birnie *et al.*, 2018; Cheng, 2022; Hassandra *et al.*, 2021; Kamińska *et al.*, 2022; Pedroli *et al.*, 2018; Schmidt *et al.*, 2021; Verkuyl *et al.*, 2018]. Por otro lado, se ha identificado un único estudio que evalúa un entorno de RA [Rohrbach *et al.*, 2019]. Además, también se ha identificado la evaluación de un entorno de RM [Rzeszewski y Orylski, 2021].

La **evaluación heurística** es un proceso sistemático de evaluación de la usabilidad del diseño de la interfaz de usuario. Los partidarios de la evaluación heurística sostienen que, mediante el uso de las pautas de usabilidad, o también llamadas principios heurísticos, para evaluar la interfaz de usuario, se pueden identificar problemas en la aplicación y evaluar el impacto de aquellos problemas. Esta técnica es considerada efectiva, fácil de aprender y utilizar, y asequible, ya que requiere la participación de solo unos pocos evaluadores para evaluar la aplicación de la técnica, regularmente los evaluadores son expertos en el área [Nielsen, 1994a]. La técnica es utilizada e informada por 8 estudios primarios [Dantes *et al.*, 2020; Doumanis y Economou, 2019; Fajar *et al.*, 2022; Kabassi *et al.*, 2018; Li *et al.*, 2019; Mitre-Ortiz *et al.*, 2022; Tromp *et al.*, 2018; Verkuyl *et al.*, 2018]. La evaluación heurística fue aplicada para evaluar la usabilidad en 2 tipos entornos de RX. Por un lado, encontramos estudios que reportan el uso de esta técnica en entornos de RV [Doumanis y Economou, 2019; Fajar *et al.*, 2022; Kabassi *et al.*, 2018; Li *et al.*, 2019; Mitre-Ortiz *et al.*, 2022; Tromp *et al.*, 2018; Verkuyl *et al.*, 2018]. Y por otro lado, también es informada en estudios de entornos de RA [Dantes *et al.*, 2020].

La **encuesta** es una técnica comúnmente utilizada para evaluar la usabilidad, permitiendo recopilar información y opiniones de una muestra de personas sobre un tema específico. Consiste en formular una serie de preguntas estructuradas, ya sean abiertas o cerradas, que son respondidas por los encuestados, lo cual proporciona datos relevantes sobre el tema en investigación [Preece *et al.*, 1994; Shneiderman, 1998]. En los estudios de usabilidad encontrados en la literatura, se utilizan distintos tipos de encuestas para obtener información. Uno de ellos es el SUS, una escala tipo Likert compuesta por 10 elementos, que simplifica la evaluación de la usabilidad en la industria y brinda una visión general de las evaluaciones subjetivas de usabilidad

[Brooke, 1996]. El SUS ha sido reportado en varios estudios primarios [De Melo *et al.*, 2019; Hassandra *et al.*, 2021; Li *et al.*, 2019; Pedroli *et al.*, 2018; Rzeszewski y Orylski, 2021; Schmidt *et al.*, 2021]. En relación a las encuestas adaptadas, se modifican y personalizan las preguntas y escalas de calificación para abordar aspectos específicos del sistema y considerar el contexto de uso, el perfil de los usuarios y los objetivos de evaluación. Esta adaptación implica ajustar el lenguaje utilizado, adaptar las escalas de calificación e incluir preguntas relevantes para el sistema en cuestión. La técnica de encuestas adaptadas es reportada en 3 estudios primarios [Birt y Vasilevski, 2021; Strada *et al.*, 2019; Verkuyl *et al.*, 2018]. Esta técnica ha sido utilizada en 8 estudios primarios para evaluar la usabilidad de entornos de RV [Birt y Vasilevski, 2021; De Melo *et al.*, 2019; Hassandra *et al.*, 2021; Li *et al.*, 2019; Pedroli *et al.*, 2018; Schmidt *et al.*, 2021; Strada *et al.*, 2019; Verkuyl *et al.*, 2018] y tan solo un estudio reporta su uso para evaluar la usabilidad de un entorno de RM [Rzeszewski y Orylski, 2021].

Según [Preece *et al.*, 1994; Shneiderman, 1998], un **experimento de usabilidad** es una técnica sistemática y controlada para evaluar la facilidad de uso y aprendizaje de un producto de software. Se lleva a cabo mediante pruebas con usuarios representativos en un entorno simulado, que permiten identificar fortalezas y debilidades. Los datos de desempeño y percepción de los usuarios se recopilan y analizan para mejorar la calidad de la interfaz y la experiencia del usuario. Los experimentos miden variables como el tiempo de respuesta, la tasa de error, la eficiencia, la satisfacción del usuario y la retención de información. El diseño del experimento debe ser riguroso y considerar factores como la selección de usuarios, el escenario de prueba y las métricas de evaluación. Los experimentos de usabilidad pueden ser realizados en diferentes etapas del ciclo de vida del software y con diferentes metodologías según los objetivos específicos del proyecto. Esta técnica fue utilizada en 3 estudios primarios para evaluar la usabilidad [Birt y Vasilevski, 2021; Li *et al.*, 2019; Rohrbach *et al.*, 2019]. De estos, dos evalúan entornos de RV [Birt y Vasilevski, 2021; Li *et al.*, 2019], mientras que solo uno evalúa un entorno de RA [Rohrbach *et al.*, 2019].

La **revisión de expertos** es una técnica de evaluación de usabilidad en la que expertos evalúan la usabilidad de un producto de software con el objetivo de identificar problemas y oportunidades de mejora. Esta técnica consiste de un proceso de inspección sistemático y metódico de un producto de software para identificar problemas de usabilidad [Shneiderman, 1998]. Por su parte, según [Preece *et al.*,

1994] es una técnica en la que expertos en usabilidad evalúan la interfaz de usuario y sugieren mejoras. [Constantine y Lockwood, 1999] la describen como una técnica de evaluación enfocada en la calidad de la solución del diseño propuesto para los requisitos y necesidades del usuario. En resumen, la revisión de expertos es una técnica clave para mejorar la usabilidad de los productos de software mediante la identificación temprana de problemas de diseño. Esta técnica ha sido reportada en 2 estudios primarios [De Melo *et al.*, 2019; Schmidt *et al.*, 2021], siendo ambos estudios de entornos de RV.

El **recorrido cognitivo**, también conocido como recorrido de tareas, es una técnica de evaluación de usabilidad que se enfoca en el análisis de cómo los usuarios interactúan con un sistema de software al realizar tareas específicas. Esta técnica implica seguir a los usuarios mientras completan una tarea y registrar su proceso de pensamiento, decisiones y acciones. Existen trabajos que coinciden en que esta técnica es útil para identificar problemas de usabilidad, evaluar la eficacia de la interfaz de usuario y mejorar la experiencia del usuario [Constantine y Lockwood, 1999; Mayhew, 1999; Preece *et al.*, 1994; Shneiderman, 1998]. Además, sugieren que esta técnica debe ser realizada por expertos en usabilidad y que el diseño de la tarea y la selección de usuarios deben ser cuidadosamente considerados. Se reporta el uso de la técnica en 2 estudios primarios [Doumanis y Economou, 2019; Tromp *et al.*, 2018]. En el estudio [Tromp *et al.*, 2018], se utilizó la técnica sin adaptaciones para evaluar un entorno de RV. Mientras que, en el estudio [Doumanis y Economou, 2019], se reporta el uso de la técnica adaptada por los autores, que también evaluó un entorno de RV.

La **retroalimentación del usuario** a través de comentarios, según [Mayhew, 1999; Nielsen, 1994b], se refiere a la información que los usuarios proporcionan acerca de su experiencia de uso de un producto de software. Esta información puede incluir comentarios sobre problemas de usabilidad, sugerencias de mejora o cualquier otro tipo de retroalimentación que el usuario considere relevante para mejorar la calidad del producto. La retroalimentación del usuario es de gran importancia para los diseñadores y desarrolladores de software, ya que permite mejorar la calidad del producto y asegurar que cumpla con las necesidades y expectativas de los usuarios finales. Esta técnica ha sido reportada por 2 estudios primarios [Birnie *et al.*, 2018; Ogunseiju *et al.*, 2022]. El primer estudio [Birnie *et al.*, 2018] reporta la evaluación

de un entorno de RV. Mientras que, en el segundo estudio [Ogunseiju *et al.*, 2022], se evalúa un entorno de RM.

De acuerdo con [Constantine y Lockwood, 1999; Nielsen, 1994b; Preece *et al.*, 1994], el **pensamiento en voz alta** es una técnica que se utiliza para evaluar la usabilidad de productos de software. Esta técnica implica pedirle a los usuarios que expresen en voz alta sus pensamientos y comentarios mientras realizan tareas específicas en el producto. Al escuchar los pensamientos de los usuarios, los evaluadores pueden obtener información valiosa sobre los problemas de usabilidad que puedan estar enfrentando los usuarios mientras utilizan el producto. La información obtenida de esta técnica puede ser utilizada para identificar problemas de diseño y mejorar la experiencia de usuario. El pensamiento en voz alta es una técnica comúnmente utilizada en las pruebas de usabilidad y puede llevarse a cabo tanto en laboratorios especializados como en entornos naturales. Esta técnica ha sido informada por 4 estudios primarios [Besoain *et al.*, 2021; Ogunseiju *et al.*, 2022; Tromp *et al.*, 2018; Verkuyl *et al.*, 2018]. En tres de los estudios primarios se evalúa un entorno de RV [Besoain *et al.*, 2021; Tromp *et al.*, 2018; Verkuyl *et al.*, 2018]. Mientras que, solo en uno de los estudios primarios se evalúa un entorno de RM [Ogunseiju *et al.*, 2022].

Según los autores [Hix y Hartson, 1993; Nielsen, 1994b; Preece *et al.*, 1994], las **grabaciones de pruebas de usabilidad** son una técnica utilizada en la evaluación de la usabilidad de productos de software. Esta técnica implica el registro de audio y video de las interacciones entre los usuarios y el producto durante las pruebas de usabilidad. Las grabaciones permiten capturar de manera precisa y objetiva los comentarios, acciones y reacciones de los usuarios mientras utilizan el producto. Estas grabaciones se utilizan posteriormente para analizar detalladamente el comportamiento de los usuarios, identificar problemas de usabilidad y realizar mejoras en el diseño y la experiencia de usuario. Las grabaciones de pruebas de usabilidad proporcionan una fuente de datos rica, así como, permiten a los evaluadores y diseñadores obtener una comprensión profunda de las fortalezas y debilidades del producto desde la perspectiva del usuario. Esta técnica se reporta en 2 estudios primarios [Kamińska *et al.*, 2022; Ogunseiju *et al.*, 2022]. En el primer estudio, se realiza una evaluación de usabilidad a entornos de RV [Kamińska *et al.*, 2022]. Mientras que, en el segundo estudio se evalúa la usabilidad a un entorno de RM [Ogunseiju *et al.*, 2022].

Los **cuestionarios** implican la recopilación de información a través de preguntas estructuradas y son valiosos para obtener datos cuantitativos sobre la satisfacción

del usuario, percepciones y preferencias. Estos cuestionarios suelen incluir escalas de calificación, preguntas de opción múltiple y preguntas abiertas que permiten a los usuarios expresar sus opiniones sobre diferentes aspectos del producto. Además, se enfatiza que los cuestionarios son versátiles y se pueden utilizar en diversas etapas del ciclo de vida del diseño, siendo útiles tanto en la fase de investigación para comprender las necesidades y expectativas de los usuarios, como en la fase de evaluación para medir la eficacia y eficiencia del producto [Mayhew, 1999; Nielsen y Molich, 1990]. Además, ambos autores destacan la importancia de los cuestionarios como técnica de evaluación de usabilidad. Tres estudios primarios informan sobre la aplicación de esta técnica [Delamarre *et al.*, 2020; Kamińska *et al.*, 2022; Ogunseju *et al.*, 2022]. Al igual que la técnica anterior, se evalúan entornos de RV [Delamarre *et al.*, 2020; Kamińska *et al.*, 2022] y RM [Ogunseju *et al.*, 2022].

En el ámbito de la usabilidad, la técnica **Personas** desempeña un papel fundamental como herramienta para ayudar a crear interfaces centradas en el usuario. Esta técnica consiste en la creación de personajes ficticios que representan diferentes tipos de usuarios, basados en una exhaustiva investigación y análisis de las necesidades, como de los comportamientos de los usuarios reales. Al emplear la técnica Personas, los diseñadores pueden obtener una comprensión más profunda de las necesidades y objetivos de los usuarios, lo que les permite diseñar interfaces que satisfagan dichas necesidades. Además, la técnica Personas es útil para identificar posibles problemas de usabilidad al resaltar las áreas en las que diferentes usuarios pueden tener distintas necesidades o preferencias. Esto proporciona una visión más integral y completa de los posibles escenarios de uso, lo cual a su vez ayuda a los diseñadores a anticipar y abordar de manera proactiva posibles obstáculos o dificultades que podrían surgir durante la interacción con la interfaz [Cooper *et al.*, 2007]. La técnica Personas ha sido empleada por tan solo un estudio primario [Tromp *et al.*, 2018] el cual, evalúa un entorno de RV.

El **seguimiento ocular** es una valiosa técnica en el proceso de evaluación de usabilidad, ya que nos brinda información relevante sobre la interacción de los usuarios con una interfaz. Nos permite comprender qué están viendo y qué están pasando por alto cuando enfrentan dificultades persistentes en una tarea específica o si ignoran elementos importantes. Además, esta técnica nos ofrece la posibilidad de grabar los recorridos visuales de varios usuarios y superponerlos para crear mapas de calor que visualizan las áreas de la interfaz de usuario que son más frecuentemente

vistas. Estos mapas de calor resultan muy útiles para identificar las secciones de la interfaz que generan mayor interés y atención, así como aquellas que podrían pasar desapercibidas [O’Grady y O’Grady, 2017]. Cabe destacar que esta técnica ha sido utilizada en un solo estudio primario, realizado por [Ogunseiju *et al.*, 2022], donde evaluaron un entorno de RM.

Las **métricas de usabilidad** son medidas utilizadas para evaluar la calidad de la experiencia del usuario al interactuar con un software [Constantine y Lockwood, 1999]. Estas métricas se enfocan en aspectos clave como la eficiencia, eficacia, satisfacción del usuario y facilidad de aprendizaje. Los autores destacan la importancia de diseñar meticulosamente estas métricas, adaptándolas a las necesidades del proyecto y los usuarios, para asegurar que sean relevantes, medibles y basadas en criterios claros y precisos. Asimismo, es esencial que las métricas sean comparables a lo largo del tiempo y proporcionen una base sólida para la toma de decisiones en el proceso de diseño y desarrollo. Para medir estas métricas, se pueden emplear diversas técnicas, como pruebas de usabilidad, encuestas y análisis de datos de uso del software, lo cual permite identificar problemas de usabilidad y mejorar la calidad de la experiencia del usuario a través de iteraciones continuas en el diseño e implementación del producto. La técnica es reportada tan solo por un estudio primario [Alpala *et al.*, 2022] evaluando un entorno de RV.

Finalmente, se encuentra la técnica **observación directa** que se define como una técnica de recopilación de datos en el campo de la IPC. Consiste en observar directamente a los usuarios mientras interactúan con un sistema o interfaz, registrando su comportamiento, acciones y reacciones. Esta técnica proporciona datos objetivos sobre cómo los usuarios utilizan el sistema, identifica problemas de usabilidad, comprende sus necesidades y obtiene información detallada sobre su experiencia. Además, permite a los investigadores obtener una visión más precisa y contextualizada de la interacción usuario-sistema [Preece *et al.*, 1994]. Es importante mencionar que la técnica de observación directa sólo fue reportada por un estudio primario [Be-soain *et al.*, 2021], el cual evalúa un entorno de RV. En la Tabla 4.1 se presenta un resumen de las técnicas identificadas en el EMS que se están utilizando para evaluar la usabilidad de entornos de RX. Para cada técnica se especifica el estudio primario que reporta su uso y el tipo de entorno evaluado.

Tabla 4.1: Técnicas de evaluación de usabilidad

Técnicas de usabilidad	Estudios Primarios	Tipo de entorno evaluado
Entrevista	[Birnie <i>et al.</i> , 2018; Cheng, 2022; Hassandra <i>et al.</i> , 2021; Kamińska <i>et al.</i> , 2022; Pedroli <i>et al.</i> , 2018; Schmidt <i>et al.</i> , 2021; Verkuyl <i>et al.</i> , 2018]	RV
	[Rohrbach <i>et al.</i> , 2019]	RA
	[Rzeszewski y Orylski, 2021]	RM
Encuesta	[Birt y Vasilevski, 2021; De Melo <i>et al.</i> , 2019; Hassandra <i>et al.</i> , 2021; Li <i>et al.</i> , 2019; Pedroli <i>et al.</i> , 2018; Schmidt <i>et al.</i> , 2021; Strada <i>et al.</i> , 2019; Verkuyl <i>et al.</i> , 2018]	RV
	[Rzeszewski y Orylski, 2021]	RM
Evaluación heurística	[Doumanis y Economou, 2019; Fajar <i>et al.</i> , 2022; Kabassi <i>et al.</i> , 2018; Li <i>et al.</i> , 2019; Mitre-Ortiz <i>et al.</i> , 2022; Tromp <i>et al.</i> , 2018; Verkuyl <i>et al.</i> , 2018]	RV
	[Dantes <i>et al.</i> , 2020]	RA
Experimento de usabilidad	[Birt y Vasilevski, 2021; Li <i>et al.</i> , 2019]	RV
	[Rohrbach <i>et al.</i> , 2019]	RA
Revisión de expertos	[De Melo <i>et al.</i> , 2019; Schmidt <i>et al.</i> , 2021]	RV
Retroalimentación del usuario	[Birnie <i>et al.</i> , 2018]	RV
	[Ogunseiju <i>et al.</i> , 2022]	RM
Pensamiento en voz alta	[Besoain <i>et al.</i> , 2021; Tromp <i>et al.</i> , 2018; Verkuyl <i>et al.</i> , 2018]	RV
	[Ogunseiju <i>et al.</i> , 2022]	RM
Grabaciones de pruebas de usabilidad	[Kamińska <i>et al.</i> , 2022]	RV
	[Ogunseiju <i>et al.</i> , 2022]	RM
Cuestionarios	[Delamarre <i>et al.</i> , 2020; Kamińska <i>et al.</i> , 2022]	RV
	[Ogunseiju <i>et al.</i> , 2022]	RA
Recorrido cognitivo	[Doumanis y Economou, 2019; Tromp <i>et al.</i> , 2018]	RV
Personas	[Tromp <i>et al.</i> , 2018]	RV
Seguimiento ocular	[Ogunseiju <i>et al.</i> , 2022]	RM
Métricas de usabilidad	[Alpala <i>et al.</i> , 2022]	RV
Observación directa	[Besoain <i>et al.</i> , 2021]	RV

4.2. Entornos de Realidad Extendida Evaluados

En esta sección, se responde a la tercera pregunta de investigación *¿Cuáles son los entornos de realidad extendida que se están evaluando actualmente?* Luego de realizar un análisis de los estudios primarios, se identificaron tres tipos de entornos de RX: RV, RA y RM. Un total de 25 estudios primarios reportan la evaluación de usabilidad de sistemas software de RV [Alpala *et al.*, 2022; Barricelli *et al.*, 2018; Besoain *et al.*, 2021; Birnie *et al.*, 2018; Birt y Vasilevski, 2021; Cheng, 2022; De Melo *et al.*, 2019; Delamarre *et al.*, 2020; Doumanis y Economou, 2019; Fajar *et al.*, 2022; Granić *et al.*, 2020; Harms, 2019; Hassandra *et al.*, 2021; He *et al.*, 2018; Jangid y Kongsilp, 2021; Kabassi *et al.*, 2018; Kamińska *et al.*, 2022; Li *et al.*, 2019; Liu *et al.*, 2018; Mitre-Ortiz *et al.*, 2022; Pedroli *et al.*, 2018; Schmidt *et al.*, 2021;

Thalmann *et al.*, 2021; Tromp *et al.*, 2018; Verkuyl *et al.*, 2018]. La RV, en general, se define cómo aquella realidad en la que el participante está totalmente inmerso en un mundo ficticio donde pueden interactuar con dicha realidad. Ese mundo puede simular las propiedades de ciertos entornos del mundo real, ya sean existentes o ficticios. Además para poder sumergirse en dicha RV, son necesarios dispositivos de visualización, los cuales funcionan como barreras para mantener al usuario dentro de este mundo virtual [Çöltekin *et al.*, 2020; Milgram y Kishino, 1994].

Solamente tres estudios primarios [Dantes *et al.*, 2020; Rohrbach *et al.*, 2019; Strada *et al.*, 2019] reportan la evaluación de usabilidad de sistemas software de RA. Se puede decir que la RA es una variación de la RV. La definición general de RA es una tecnología que une el espacio y los objetos reales con objetos generados por computador (virtuales), los cuales son superpuestos en el mundo real en tiempo real. La principal diferencia entre RV y RA es que la RA no sustituye por completo el mundo real [Azuma, 1997; Rzeszewski y Orylski, 2021].

Con respecto a la RM, solo dos estudios primarios [Ogunseiju *et al.*, 2022; Rzeszewski y Orylski, 2021] reportan la evaluación de usabilidad de sistemas software con esta tecnología. La RM une objetos del tipo virtual con otros objetos del mundo real. La principal diferencia entre la RM y la RA es que, en el caso de esta última, los objetos virtuales no sólo se superponen al mundo real, sino que también se puede interaccionar con ellos. Los objetos digitales de RM están ligados al espacio real, por ejemplo, un holograma puede verse superpuesto a un objeto concreto, por el contrario, en la RA, un holograma de un objeto específico puede colocarse sobre cualquier superficie plana [Rzeszewski y Orylski, 2021].

4.3. Principales Problemas y Desafíos en la Aplicación de Técnicas para Evaluar la Usabilidad de Entornos Virtuales

En esta sección, se da respuesta a la segunda pregunta de investigación *¿Cuáles son los principales problemas y desafíos en la aplicación técnicas para evaluar la usabilidad de los entornos virtuales?* De una inspección exhaustiva, se encontró que solo cuatro de los estudios primarios [Delamarre *et al.*, 2020; Granić *et al.*, 2020; Kabassi *et al.*, 2018; Verkuyl *et al.*, 2018] reportan los resultados (es decir, si la

aplicación de la técnica fue exitosa o por el contrario se presentaron problemas o retos) de aplicar las técnicas de usabilidad o *frameworks* propuestos por los autores.

Para comenzar, es fundamental mencionar que existen problemas de usabilidad relacionados con la muestra [Granić *et al.*, 2020; Verkuyl *et al.*, 2018]. En el primer caso, el muestreo se realiza por conveniencia, es decir, un método no probabilístico ni aleatorio, donde el investigador escoge la muestra por proximidad y sin considerar si es o no representativa. En el segundo caso, otro reto relacionado con la muestra, son las altas tasas de abandono [Verkuyl *et al.*, 2018]. Además, la realización de la evaluación de usabilidad en un contexto específico y con participantes expertos contribuye a restringir la generalización de los resultados [Granić *et al.*, 2020]. Luego están los problemas y retos relacionados directamente con la aplicación de las técnicas de evaluación de usabilidad [Delamarre *et al.*, 2020; Kabassi *et al.*, 2018]. En el caso de la técnica de evaluación heurística [Kabassi *et al.*, 2018], el problema principal radica en la subjetividad de dicha evaluación, ya que, no deja de ser una opinión, donde los expertos que realizan la evaluación heurística entregan sus impresiones por sobre una opinión fundamentada. La técnica de pensamiento en voz alta [Delamarre *et al.*, 2020] presenta un problema relacionado con la sensación de presencia en los entornos virtuales. Cuando los usuarios hablan en voz alta sobre sus acciones y pensamientos durante la interacción con el sistema, pueden romper la sensación de estar realmente presentes en el entorno virtual. Esta situación puede constituir una limitación significativa en la evaluación de la usabilidad de dichos entornos virtuales. Otro problema que se visibiliza en el estudio de [Verkuyl *et al.*, 2018], es debido a limitaciones de tiempo y recursos económicos, únicamente se pudo llevar a cabo una única ronda de evaluaciones, lo que impidió la realización de una segunda prueba posterior a las modificaciones realizadas en el juego.

Los problemas relacionados a la usabilidad de los diferentes entornos de realidad extendida son reportados por 7 estudios primarios. En el primer estudio [Liu *et al.*, 2018], se identificaron varios problemas de usabilidad relacionados con la planificación de misiones de drones en realidad virtual. Los usuarios reportaron confusión y ambigüedad en la interfaz, una curva de aprendizaje pronunciada, sobrecarga cognitiva debido a controles excesivos, falta de retroalimentación y preocupaciones de seguridad. Para abordar estos problemas, se recomendaron mejoras como tutoriales para nuevos usuarios y la simplificación de la interfaz. En el segundo estudio [Doumanis y Economou, 2019], se destacaron problemas adicionales en la usabilidad

de prototipos de realidad virtual. Se mencionaron dificultades en la navegación del avatar y el reconocimiento de atajos de teclado, así como problemas relacionados con el diseño estético y la interacción en la interfaz de usuario. Este estudio subraya la importancia de abordar los problemas de usabilidad tempranamente en el proceso de diseño. El tercer estudio [Granić *et al.*, 2020] evaluó un entorno virtual de aprendizaje, identificando problemas técnicos, de diseño de interfaz, experiencia de aprendizaje deficiente y aburrimiento con el entorno. Se sugirió que el diseño de la interfaz necesita mejoras para brindar experiencias de aprendizaje en línea más efectivas. En el cuarto estudio [Dantes *et al.*, 2020], se evaluó una aplicación de realidad aumentada. Se mencionaron problemas como tamaños inconsistentes de marcadores, posicionamiento de objetos, problemas de percepción, navegación y manipulación, así como problemas de interacción. La aplicación requería ajustes significativos para mejorar su usabilidad. El quinto estudio [Delamarre *et al.*, 2020] el uso de punteros láser en un entorno de escritorio de realidad virtual. También discutió limitaciones de los métodos tradicionales de evaluación de usabilidad en entornos virtuales y destacó la importancia de un enfoque tutorial autónomo. El sexto estudio señaló problemas como la falta de claridad sobre la entrada al entorno inmersivo y la confusión en torno al botón “Enter RV”. Además, mencionó obstáculos visuales y requisitos específicos para lograr niveles satisfactorios de usabilidad en WebXR [Rzeszewski y Orylski, 2021]. En el último estudio [Cheng, 2022], se identificó problemas como la falta de interacción funcional en las experiencias de campo virtual, dificultades para obtener información en un idioma específico y problemas de hardware y red. Estos problemas afectaron la efectividad del aprendizaje en entornos virtuales.

4.4. Cómo se están Aplicando las Técnicas para la Evaluación de la Usabilidad de Entornos Virtuales

En esta sección, se llevará a cabo una exploración exhaustiva de las técnicas utilizadas para evaluar la usabilidad de entornos virtuales. Se orientará en el análisis de los enfoques y metodologías actualmente empleados para medir parámetros de calidad, como la eficacia, eficiencia y satisfacción del usuario. Se ha identificado que la evaluación de usabilidad de los entornos de RX se está realizando de 5 maneras

diferentes: (i) aplicando técnicas de evaluación de la usabilidad tradicionales según lo prescrito por la teoría; (ii) adaptando técnicas de evaluación de usabilidad tradicionales; (iii) combinando técnicas de evaluación de usabilidad; (iv) proponiendo y aplicando *frameworks*; (v) proponiendo y aplicando nuevas técnicas de evaluación de usabilidad. A continuación, se describirán los puntos anteriores (a excepción del primero que fue descrito en la Sección 4.1). El objetivo principal de esta sección es brindar una visión integral y detallada de cómo se aborda actualmente la evaluación de la usabilidad en entornos virtuales.

En primer lugar, con respecto a las **técnicas de evaluación de usabilidad adaptadas**, se identificaron dos: (i) encuesta [Birt y Vasilevski, 2021; Verkuyl *et al.*, 2018], y (ii) recorrido cognitivo [Doumanis y Economou, 2019]. La encuesta adaptada la encontramos en los estudios mencionados anteriormente. En el primer estudio [Verkuyl *et al.*, 2018], se utilizan varias técnicas de evaluación de usabilidad: pensamiento en voz alta, encuesta adaptada y finalmente una entrevista. Nos centraremos en explicar como fue aplicada la encuesta, como el segundo medio de recopilación de datos para evaluar la usabilidad de un juego virtual de simulación (VGS, por sus siglas en inglés de *Virtual Gaming Simulation*). Esta encuesta constaba de cuatro preguntas demográficas y 18 preguntas que evaluaban la facilidad de uso y utilidad del juego utilizando una escala Likert de cinco puntos. Para desarrollar estas preguntas, el equipo de investigación adaptó las preguntas de la encuesta con base en pruebas de usabilidad anteriores de VGS, el modelo de aceptación de tecnología (TAM, por sus siglas en inglés de *Technology Acceptance Model*) y la literatura de usabilidad. De esta manera, utilizaron investigaciones y marcos teóricos existentes para asegurarse de que las preguntas de la encuesta estuvieran midiendo aspectos relevantes de la usabilidad. En el segundo estudio [Birt y Vasilevski, 2021], la encuesta adaptada es utilizada para evaluar la usabilidad y la experiencia de la RV inmersiva (IVR, por sus siglas en inglés de *Immersive Virtual Reality*). La encuesta adaptada es una medida autoinformada que incluye preguntas adaptadas del líder en usabilidad de software Nelson/Norman Group y del Modelo de Aceptación de Hardware de RV (VR-HAM, por sus siglas en inglés de *Virtual Reality Hardware Acceptance*) de Manis y Choi (2019). Aunque no se detallan los elementos específicos de la encuesta, se identifican tres dimensiones de usabilidad de la IVR móvil: utilidad, compromiso y experiencia. Todos los elementos fueron evaluados en una escala de Likert de cinco puntos, desde “fuertemente en desacuerdo” hasta “fuertemente de

acuerdo”. La encuesta se administró utilizando la plataforma en línea *Qualtrics*, en concordancia con los principios éticos, mediante un código QR con un identificador único.

La técnica recorrido cognitivo adaptada es reportada por un solo estudio [Domanis y Economou, 2019]. La técnica inicia con un análisis detallado de las tareas necesarias, donde el experimentador establece una secuencia de acciones que el usuario debe realizar para completar la tarea, así como las respuestas del sistema a esas acciones. Los evaluadores siguen estos pasos y plantean cuatro preguntas para cada paso de las tareas asignadas. Aunque las respuestas a las preguntas son de formato binario (Sí/No), también se les pide a los evaluadores que comenten sobre su respuesta preferida. Además, se les solicita que indiquen en una escala del 0 % al 100 % la probabilidad de que los usuarios tengan dificultades para realizar correctamente cada acción, según los requisitos establecidos. Para guiar a los evaluadores sobre el tipo de información esperada, se adaptó la técnica al proporcionar explicaciones de texto adicionales debajo de cada pregunta. También se les asignaron personajes que representaban a diferentes usuarios de los prototipos de RV, y se integraron tareas en casos de uso que reflejaban los requisitos de cada escenario educativo. El diseño de los personajes se basó en los requisitos iniciales de los usuarios recopilados durante las etapas iniciales del proyecto de participación real y virtual en un entorno de inmersión realista (REVERIE, por sus siglas en inglés de *Real and Virtual Engagement In Realistic Immersive Environment*). Se recopilaron datos cuantitativos de 277 usuarios a través de una encuesta en línea que abordaba varios aspectos del sistema REVERIE, y también se obtuvieron datos cualitativos de posibles usuarios mediante dos inspecciones informales de usabilidad. La técnica de recorrido cognitivo adaptada en el estudio involucró a dos evaluadores que revisaron las tareas desde la perspectiva de un estudiante, y a un experto que lo hizo desde la perspectiva de un profesor. Se les proporcionó a los evaluadores un formulario estándar de recorrido cognitivo que enumeraba las tareas que debían revisar y las herramientas que debían utilizar para cada tarea (por ejemplo, prototipo de software, guiones gráficos o navegador de internet). Al inicio de cada sesión, se capacitó a los evaluadores sobre cómo utilizar la técnica, una sesión que duró 10 minutos e incluyó instrucciones y preguntas y respuestas. En total, los evaluadores analizaron 36 tareas agrupadas en cuatro categorías. Se solicitó a los evaluadores que revisaran las tareas desde la perspectiva de un estudiante, lo que implicaba imaginar que eran estudiantes utili-

zando el prototipo de RV por primera vez. También se les proporcionó un conjunto de escenarios que describían situaciones típicas en las que los estudiantes podrían utilizar el prototipo. Para cada tarea, los evaluadores tenían que responder un conjunto de preguntas centradas en tres aspectos principales: formulación del objetivo, evaluación de la acción y evaluación del sistema. Además, se les pidió que calificaran cada tarea en una escala del 1 al 5 según su dificultad al utilizar el prototipo de RV. Una vez completadas todas las tareas, se solicitó a los evaluadores que proporcionaran comentarios sobre su experiencia utilizando la técnica de recorrido cognitivo y sugirieran mejoras para futuras evaluaciones.

En segundo lugar, se encontraron diversas **combinaciones de técnicas** que han sido utilizadas en conjunto para obtener una evaluación más completa. En el presente trabajo de titulación, se reconocieron 12 combinaciones diferentes. En primer lugar, la técnica entrevista se reporta combinada de cinco maneras diferentes: (i) aplicada en conjunto con la técnica retroalimentación del usuario [Birnie *et al.*, 2018]; (ii) en conjunto con evaluación heurística, encuesta y pensamiento en voz alta [Verkuyl *et al.*, 2018]; (iii) es utilizada junto a la encuesta [Schmidt *et al.*, 2021]; (iv) se utiliza junto con un experimento de usabilidad [Rohrbach *et al.*, 2019]; y (v) utilizada en conjunto a un cuestionario [Kamińska *et al.*, 2022].

La evaluación heurística se combina de dos formas: (i) En conjunto con pensamiento en voz alta, recorrido cognitivo y Personas [Tromp *et al.*, 2018]; (ii) se aplica junto con la técnica recorrido cognitivo [Doumanis y Economou, 2019]. La encuesta se utiliza en combinación con otras técnicas de cuatro maneras: (i) se emplea junto con la técnica entrevista y revisión de expertos [Schmidt *et al.*, 2021]; (ii) se aplica en conjunto a la revisión de expertos [De Melo *et al.*, 2019]; (iii) se reporta en conjunto con la evaluación heurística y experimento de usabilidad [Liu *et al.*, 2018]; (iv) se aplica junto a un experimento de usabilidad [Birt y Vasilevski, 2021]. Por último, la retroalimentación del usuario se reporta en conjunto con pensamiento en voz alta, cuestionario, seguimiento ocular y grabaciones de las pruebas de usabilidad [Ogunseiju *et al.*, 2022]. La información anterior se sintetiza en la Tabla 4.2.

En tercer lugar, se identifican cinco estudios donde los autores han propuesto **framework** para evaluar la usabilidad en entornos de RV [Barricelli *et al.*, 2018; He *et al.*, 2018; Jangid y Kongsilp, 2021; Liu *et al.*, 2018; Thalmann *et al.*, 2021]. El primer estudio [He *et al.*, 2018], se enfoca en la evaluación de la usabilidad de entornos de RV basados en representaciones artísticas. La eficacia de estos sistemas

Tabla 4.2: Combinación de las técnicas más usadas

Técnicas de usabilidad	Técnicas usadas en conjunto	Tipo de entorno evaluado
Entrevista	Retroalimentación del usuario [Birnie <i>et al.</i> , 2018]	RV
	Evaluación heurística, encuesta y pensamiento en voz alta [Verkuyl <i>et al.</i> , 2018]	RV
	Encuesta [Schmidt <i>et al.</i> , 2021]	RV
	Experimento de usabilidad [Rohrbach <i>et al.</i> , 2019]	RA
	Cuestionario [Kamińska <i>et al.</i> , 2022]	RV
Encuestas	Entrevista y la revisión de expertos [Schmidt <i>et al.</i> , 2021]	RV
	Revisión de expertos [De Melo <i>et al.</i> , 2019]	RV
	Evaluación heurística y un experimento de usabilidad [Liu <i>et al.</i> , 2018]	RV
	Experimento de usabilidad [Birt y Vasilevski, 2021]	RV
Evaluación heurística	Pensamiento en voz alta, el recorrido cognitivo y personas [Tromp <i>et al.</i> , 2018]	RV
	Recorrido cognitivo [Doumanis y Economou, 2019]	RV
Retroalimentación del usuario	Pensamiento en voz alta, un cuestionario, el seguimiento ocular y grabaciones de las pruebas de usabilidad [Ogunseiju <i>et al.</i> , 2022]	RM

depende en gran medida de su capacidad para evocar una respuesta emocional similar. Para ello, se propone un *framework* de usabilidad de dos niveles. El primer nivel, se centra en la presencia del sistema de RV, que se divide en tres factores: la ilusión de lugar, la ilusión de verosimilitud y la conciencia de copresencia del público. Estos factores dependen de aspectos de hardware y la respuesta emocional del espectador hacia audiencias virtuales. El segundo nivel, aborda el compromiso del público, el recuerdo de la información y el deseo de ver el espectáculo en directo. Estos factores representan la respuesta del espectador al contenido de la representación en el entorno virtual. El compromiso del público no requiere necesariamente un alto nivel de inmersión, mientras que el recuerdo de la información y el deseo de ver la representación en directo influyen en la experiencia del espectador. El objetivo es evaluar la usabilidad de la aplicación de RV en comparación con el vídeo convencional y promover la asistencia a representaciones en directo.

En el segundo trabajo [Barricelli *et al.*, 2018], se realiza un estudio piloto destinado a validar el *framework* Semiótico de RV, desarrollando un protocolo que combinaba métodos de evaluación semiótica y cognitiva con el propósito de medir la usabilidad y la experiencia del usuario en aplicaciones de RV. Para ello, se seleccionaron cuidadosamente ocho aplicaciones que representaban los diferentes aspectos del *framework* semiótico, y se llevó a cabo una prueba de usuarios con un grupo reducido de participantes. El entorno de prueba se configuró con dispositivos tales como un casco *Samsung Gear RV*, un monitor para reflejar la experiencia y

una computadora portátil para permitir el acceso a Internet. Durante la prueba, los participantes realizaron tareas específicas en cada aplicación mientras un observador registraba detalladamente sus comportamientos. Los datos recopilados se obtuvieron mediante la técnica de encuesta y cuestionario de usabilidad, utilizando un conjunto de 26 preguntas basadas en una escala Likert. Además, se empleó una combinación de la encuesta SUS y el cuestionario de usabilidad de sistemas informáticos (CSUQ, por sus siglas en inglés de *Computer System Usability Questionnaire*). Junto a esto, se llevó a cabo una evaluación de la experiencia del usuario mediante el cuestionario sobre la experiencia de los usuarios (UEQ, por sus siglas en inglés de *User Experience Questionnaire*). Cabe destacar que todas las sesiones de prueba fueron grabadas para permitir la aplicación del Método de Evaluación de Comunicabilidad (CEM, por sus siglas en inglés *Communicability Evaluation Method*), un enfoque de ingeniería semiótica destinado a evaluar la comunicabilidad de las aplicaciones de RV. En conclusión, este estudio piloto involucró la implementación del *framework* Semiótico de RV mediante pruebas de usuario con ocho aplicaciones, lo que permitió recopilar datos valiosos sobre usabilidad y experiencia del usuario a través de cuestionarios y grabaciones de las sesiones.

En el tercer estudio [Liu *et al.*, 2018], se reporta un *framework* métrico basado en las siete etapas de acción de Norman para evaluar la usabilidad en entornos de RV. El *framework* se centra en abordar cuestiones de usabilidad relacionadas con la capacidad de los usuarios para identificar las acciones disponibles, llevar a cabo tareas específicas y determinar si el sistema se encuentra en el estado deseado. Para mejorar la interfaz de la RV, se han adoptado cuatro principios de diseño propuestos por Norman. Estos principios incluyen la visibilidad del estado del sistema y las opciones de acción, la implementación de un buen modelo conceptual que asegure la coherencia en la presentación de operaciones y resultados, el establecimiento de una buena cartografía que establezca una relación clara entre las acciones y los resultados, y la provisión de retroalimentación completa y continua sobre los resultados de las acciones de los usuarios. El proceso de evaluación de usabilidad implica pruebas individuales que tienen una duración de 30 a 40 minutos, durante las cuales se recopila información sobre la experiencia previa del usuario en RV. Durante la prueba, se solicita a los participantes que expresen en voz alta sus pensamientos mientras llevan a cabo una serie de tareas asignadas e interactúan con la interfaz de usuario, mientras que al mismo tiempo se observa su desempeño y se toman notas y puntua-

ciones basadas en su rendimiento. Después de las pruebas, se les pide a los usuarios que completen dos encuestas, y todas las interacciones virtuales son grabadas para realizar comparaciones y análisis de rendimiento. Además, se realiza una entrevista con los evaluadores para obtener información cualitativa sobre los aspectos que funcionaron bien y aquellos que presentaron dificultades durante la prueba.

En el cuarto estudio [Jangid y Kongsilp, 2021], se informa sobre el *Fishtank Sandbox*, el cual, es un *framework* de software que se utiliza para estudiar la usabilidad de las técnicas de interacción en RV en pecera (FTVR, por sus siglas en inglés de *FishTank Virtual Reality*), es un término que se refiere a una forma de RV que utiliza una configuración de visualización más simple y básica en comparación con los sistemas de RV tradicionales. En FTVR, los usuarios utilizan una pantalla plana o monitor para visualizar el entorno virtual en lugar de utilizar dispositivos de visualización inmersivos como cascos o gafas de RV. Anteriormente, los investigadores construían sus propios sistemas FTVR, lo que resultaba en diferentes algoritmos y configuraciones en sus estudios. Por el contrario, el *framework* propuesto por los autores se basa en Holo-SDK, un kit de desarrollo de software (SDK, por sus siglas en inglés de *Software Development Kit*) público que permite crear aplicaciones FTVR utilizando una computadora de escritorio, una cámara web y gafas anaglifo, que son un tipo de gafas diseñadas para crear efectos tridimensionales o estereoscópicos en imágenes o videos. Esto garantiza la coherencia entre los estudios al utilizar la misma configuración en términos de detección facial, seguimiento de la cabeza y perspectiva central. El *framework* tiene dos casos de uso principales: uno para que los participantes realicen experimentos descargando un programa ejecutable, y otro para que los investigadores implementen nuevas técnicas de interacción descargando un proyecto *Unity* y compartiendo el software ejecutable. El *framework* consta de tres componentes principales: una aplicación de escritorio, un componente de interacción y una aplicación web. La aplicación de escritorio controla el procedimiento del experimento y la interfaz de usuario, mientras que el componente de interacción permite a los investigadores crear y probar sus propias técnicas de interacción. La aplicación web almacena los datos del experimento y permite a los investigadores acceder a ellos para análisis y comparaciones. Además, el *framework* utiliza la tarea de Fitts en FTVR, basada en el test de Fitts, donde los participantes deben controlar un cursor 3D para alcanzar objetivos esféricos dentro de un espacio volumétrico. Cada experimento se centra en una técnica de interacción y consta de dos ensayos

con diferentes tamaños y distancias de los objetivos. La aplicación registra el tiempo, los clics fallidos y muestra los datos al participante al finalizar.

En el quinto estudio, la evaluación de la usabilidad se llevó a cabo siguiendo un *framework* de usabilidad establecido [Thalmann *et al.*, 2021]. En este *framework*, se empleó un enfoque de métodos mixtos que combinó diferentes técnicas de evaluación de usabilidad, se utilizaron técnicas cuantitativas y cualitativas para obtener respuestas más completas a la pregunta de investigación planteada. Los participantes del estudio fueron reclutados a través de contactos locales y anuncios públicos en Suiza. El prototipo de *exergame* del proyecto VITAAL se enfocó en tres componentes de entrenamiento: fuerza, equilibrio y cognición, y presentaba minijuegos temáticos correspondientes. Los *exergames* comprenden distintos tipos de interacción con videojuegos que demandan que el jugador participe físicamente y realice movimientos para jugar. Para evaluar la usabilidad, se utilizó un protocolo de usabilidad basado en el pensamiento en voz alta, SUS y una entrevista guiada por pautas. Durante la sesión de juego, el protocolo de usabilidad buscó recopilar datos sobre las interacciones de los participantes con el prototipo de *exergame*, incluyendo sus pensamientos, sentimientos y acciones. La técnica del pensamiento en voz alta permitió capturar las reacciones inmediatas y los comentarios de los participantes mientras utilizaban el prototipo, lo que facilitó la identificación en tiempo real de problemas de usabilidad y áreas de mejora. Posteriormente a la sesión de juego, se administraron cuestionarios y se realizaron entrevistas para recopilar comentarios más detallados sobre la experiencia de los participantes al utilizar el prototipo de *exergame*.

Por último, se propusieron **nuevas técnicas para medir la usabilidad** y la experiencia del usuario reportadas por dos estudios de entornos de RV [Granić *et al.*, 2020; Harms, 2019]. En el primer estudio, el enfoque de evaluación de usabilidad en grupo basado en escenarios (ScerGUT, por sus siglas en inglés de *Scenario-based Group Usability Testing*) se utiliza para diseñar y desarrollar soluciones de entornos virtuales de aprendizaje (VLE, por sus siglas en inglés de *Virtual Learning Environments*) 3D centrados en el alumno. También evalúa el valor educativo de los VLE. A diferencia de otros métodos que utilizan una sola técnica de usabilidad, ScerGUT emplea varios métodos, como pruebas basadas en tareas con grupos de usuarios finales, cuestionarios de actitud, pruebas de memorabilidad, entrevistas y observaciones del evaluador. Este enfoque es considerado de métodos mixtos, ya que combina datos

cuantitativos y cualitativos obtenidos de diferentes métodos de prueba. Se identificaron problemas en la evaluación experimental debido a variaciones en los escenarios de trabajo y métodos de evaluación. Las medidas cuantitativas utilizadas incluyen conocimiento adquirido, memorización y satisfacción de los usuarios, mientras que las medidas cualitativas abarcan la demografía, comentarios subjetivos de los usuarios y problemas identificados por los evaluadores. Se utilizan varios instrumentos de medición, como cuestionarios, pruebas de memoria, entrevistas y el cuaderno del evaluador para recopilar y analizar los datos [Granić *et al.*, 2020].

En el segundo estudio, se reporta el método automatizado propuesto por [Harms, 2019] para evaluar la usabilidad de las aplicaciones de RV. El enfoque consta de tres pasos principales: registro de las acciones del usuario, generación de árboles de tareas y detección de problemas de usabilidad. Este método se ha aplicado previamente a aplicaciones de escritorio y sitios web, pero ahora se traslada a la RV. En el primer paso, se registran las acciones del usuario en la RV. Dado que las acciones en la RV son diferentes de las interfaces gráficas tradicionales, se desarrolló un mecanismo de registro específico. Se interceptan y registran los eventos generados por las acciones del usuario, como agarrar objetos o mover la cabeza. Estos eventos se almacenan en un archivo de registro legible por humanos. El registro de eventos se realiza mediante la ampliación del código fuente de la RV para capturar los eventos relevantes. Se utiliza el grafo de escena de la RV para identificar y acceder a los objetos interactivos. Los eventos se registran en función de la interacción del usuario con estos objetos. Los eventos se envían a un servidor central para su almacenamiento y posterior procesamiento. En el segundo paso, se generan árboles de tareas a partir de una lista de eventos registrados utilizando un algoritmo que detecta iteraciones y secuencias en las listas de eventos. Las iteraciones representan eventos repetitivos, mientras que las secuencias representan sublistas de eventos repetitivos. Estos árboles de tareas representan el comportamiento del usuario registrado y son útiles para el análisis del uso del software. El tercer paso implica la detección de problemas de usabilidad mediante la identificación de “olores de usabilidad”. Se buscan comportamientos del usuario que indican problemas de usabilidad subyacentes. Se han definido varios olores de usabilidad, como tareas importantes y acciones ineficientes requeridas. Estos olores se detectan mediante el análisis de los árboles de tareas generados.

Capítulo 5

Discusión y Amenazas a la Validez

En este capítulo, se discuten los resultados derivados del análisis de los estudios primarios identificados durante el estudio exploratorio. Además, se informan las amenazas a la validez que surgieron durante la búsqueda de literatura.

5.1. Discusión

La discusión de los resultados obtenidos en esta investigación permite examinar los entornos de RX más utilizados, centrándose principalmente en la RV, y también identificar aquellos menos utilizados, como la RA y la RM. Además, se exploran las técnicas de evaluación de usabilidad más frecuentemente empleadas en estos entornos, analizando cómo se adaptan, combinan y crean *frameworks* y métodos para ajustarse a las particularidades de cada entorno. Asimismo, se abordan los desafíos que supone aplicar técnicas de evaluación de usabilidad en los entornos de RX. Este análisis ofrece una visión clara de las prácticas comunes en la evaluación de usabilidad en la RX y aborda los desafíos y retos que enfrentan los investigadores al aplicar ciertas técnicas de evaluación de usabilidad en este contexto, considerando las peculiaridades y limitaciones propias de los entornos de RX. Finalmente, se establece un panorama completo que contribuye a comprender el estado actual de la evaluación de usabilidad en la RX y ofrece perspectivas valiosas para futuras investigaciones en este ámbito. En la Figura 5.1, se resumen los resultados de los puntos mencionados anteriormente.

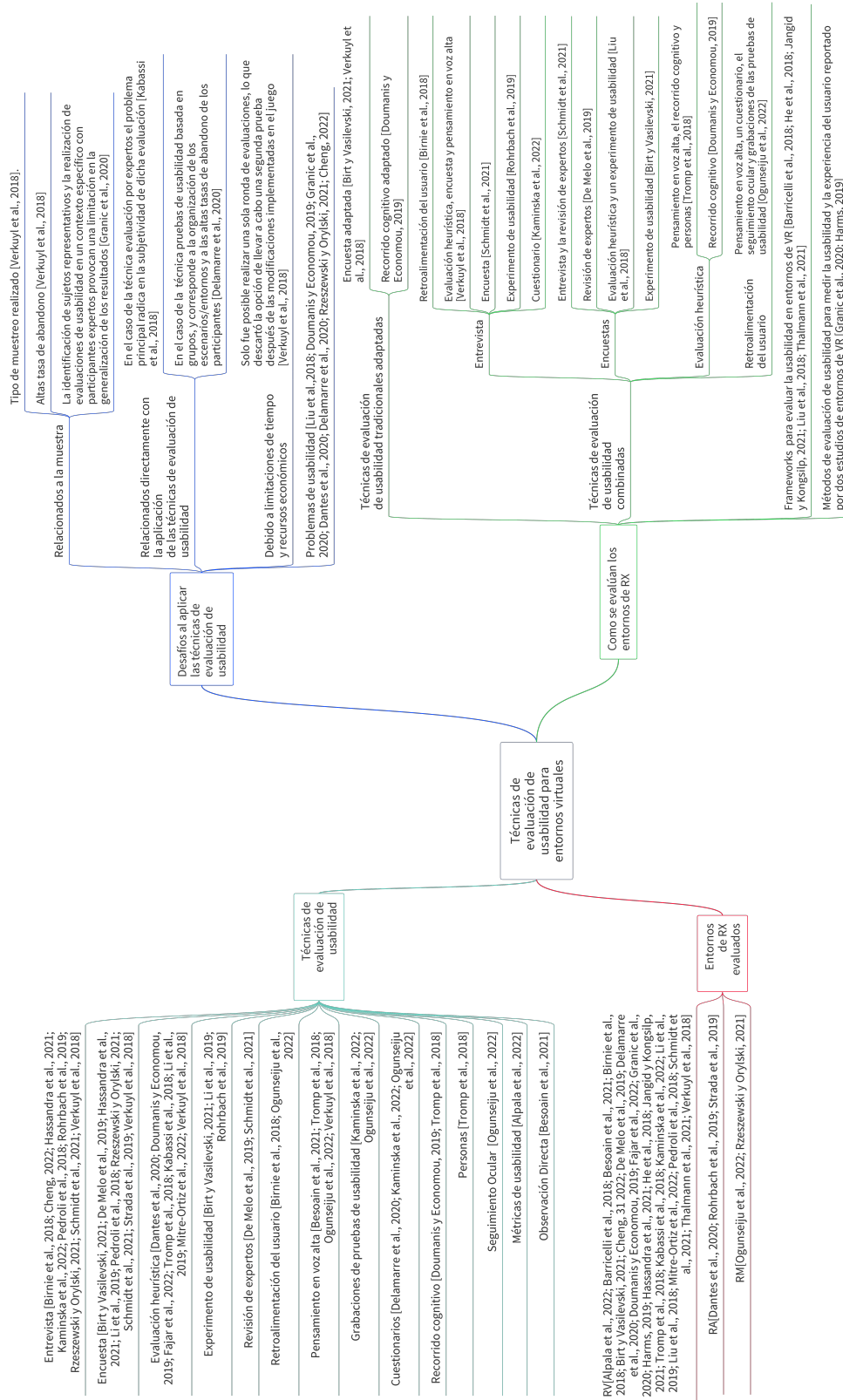


Figura 5.1: Aspectos importantes de la investigación

La evaluación de la usabilidad en entornos de RX ha sido objeto de numerosos estudios en la literatura. En el primer apartado de esta discusión, se observa que existen ciertos tipos de entornos de RX que son evaluados con mayor frecuencia que otros. Basándonos en el análisis de los estudios primarios identificados en el EMS, los entornos de RV son los más comúnmente evaluados debido a su capacidad inmersiva y su popularidad creciente en campos como el entretenimiento, la educación y la medicina. Por otro lado, los entornos de RA y RM también han sido evaluados, aunque en menor medida que la RV. La RA y RM ofrecen experiencias que combinan elementos virtuales y reales, lo que plantea desafíos únicos en términos de interacción y usabilidad. Estos resultados sugieren que la RV es el entorno de RX más consolidado y estudiado en términos de usabilidad, seguido por la RA y la RM.

En cuanto a las **técnicas utilizadas para evaluar la usabilidad** en los entornos de RX, se identificaron tres formas principales: las técnicas más usadas, las técnicas adaptadas y las técnicas combinadas. Las técnicas más utilizadas son la entrevista, la encuesta y la evaluación heurística. La entrevista permite recopilar información directa entregada por los usuarios o evaluadores, lo que permite obtener una visión detallada y contextualizada de las experiencias y necesidades en los sistemas de RX. La evaluación heurística se basa en principios heurísticos o pautas de usabilidad para identificar problemas en el diseño de la interfaz de usuario. La encuesta, por su parte, recopila información y opiniones de los usuarios sobre la usabilidad del sistema. Se destacó el uso del SUS y de encuestas adaptadas, ajustando las preguntas y escalas de calificación según el contexto y los objetivos de evaluación. En algunos casos, se requieren adaptaciones en las técnicas de evaluación para abordar las características específicas de la RX. Se identificaron dos técnicas de evaluación adaptadas: la encuesta y el recorrido cognitivo. Estas adaptaciones se deben a las diferencias en la interacción, la experiencia del usuario y las respuestas emocionales que surgen en entornos de RV. La necesidad de ajustar las técnicas radica en capturar de manera efectiva la usabilidad y la experiencia de los usuarios en este contexto específico, considerando las particularidades y desafíos que presentan los entornos de RX.

El uso de **combinaciones de técnicas** en la evaluación de la usabilidad en entornos virtuales es un tema de interés. Estas combinaciones permiten obtener una evaluación más completa y detallada, capturando una amplia gama de aspectos relacionados con la experiencia del usuario. Al combinar diferentes técnicas, co-

mo entrevistas, evaluación heurística, encuestas, pensamiento en voz alta, recorrido cognitivo y seguimiento ocular, se logra un enfoque multidimensional que aborda diversas perspectivas y aporta una mayor comprensión de la usabilidad en estos entornos. El uso de múltiples técnicas en conjunto proporciona una sinergia invaluable, ya que cada técnica aporta su propio conjunto de fortalezas y enfoques. Por ejemplo, las entrevistas pueden proporcionar información detallada y contextualizada sobre las experiencias y necesidades de los usuarios, mientras que la evaluación heurística permite identificar problemas potenciales en la interfaz de usuario. Las encuestas, por su parte, ofrecen la posibilidad de recopilar datos cuantitativos y obtener una visión más amplia de la satisfacción del usuario. Al combinar estas técnicas, se obtiene una visión más completa y precisa de la usabilidad de los entornos virtuales. Además, las combinaciones de técnicas permiten abordar los desafíos y problemas inherentes a la evaluación de la usabilidad en entornos virtuales. Por ejemplo, al enfrentar problemas de tamaño de muestra o muestreo no representativo, combinar técnicas como entrevistas y encuestas puede ayudar a obtener una perspectiva más amplia al recopilar datos cualitativos y cuantitativos de una muestra diversa de usuarios. Asimismo, al abordar la subjetividad en la evaluación por expertos, la combinación de técnicas como evaluación heurística y pensamiento en voz alta puede proporcionar una visión más objetiva y fundamentada.

En cuanto a los ***frameworks*** propuestos para evaluar la usabilidad en entornos de RV, estos buscan proporcionar estructuras y guías para evaluar de manera sistemática y efectiva la usabilidad de las aplicaciones de RV, teniendo en cuenta diferentes aspectos y perspectivas. Estos *frameworks* comparten un enfoque común al proponer marcos para evaluar la usabilidad en entornos de RV. Todos se centran en comprender y evaluar la experiencia del usuario en entornos de RV, considerando factores emocionales, cognitivos e interactivos. Se basan en una combinación de métodos y técnicas de evaluación, como pruebas de usuario, cuestionarios, entrevistas y registros de sesiones de prueba, para recopilar datos relevantes y obtener una visión completa y equilibrada de la usabilidad y la experiencia del usuario. Además, aplican principios de diseño y teorías relevantes para mejorar la usabilidad y la interacción en entornos de RV. Estos *frameworks* ofrecen estructuras sistemáticas que permiten una comprensión profunda de la usabilidad y la experiencia del usuario en los entornos de RV, y brindan la oportunidad de mejorar la usabilidad y la satisfacción del usuario. Los estudios mencionados proponen y aplican métodos específicos

para medir la usabilidad y la experiencia del usuario en entornos de RV. Reconocen que las características y las interacciones en la RV son diferentes de las interfaces tradicionales, por lo que requieren enfoques adaptados. Estos estudios destacan la necesidad de adaptar y aplicar métodos específicos para evaluar la usabilidad en entornos de RX, reconociendo las características únicas de esta tecnología y su impacto en la experiencia del usuario.

El **desafío de aplicar técnicas de evaluación de usabilidad** en entornos virtuales es un tema intrigante y relevante, que ofrece oportunidades para investigar, desarrollar y mejorar las prácticas actuales en la evaluación de la usabilidad en el ámbito de la RX. Este tema destaca la importancia de comprender y superar los obstáculos que surgen al evaluar la usabilidad en contextos de RX. La necesidad de abordar problemas como el tamaño de la muestra, el tipo de muestreo, la subjetividad en las evaluaciones y la selección de métricas adecuadas plantea interrogantes fascinantes sobre cómo mejorar la eficacia y la fiabilidad de las técnicas utilizadas. Además, este tema sugiere la importancia de adaptar las técnicas de evaluación de usabilidad tradicionales para que sean aplicables en entornos virtuales, teniendo en cuenta las particularidades de estos contextos y las limitaciones que pueden surgir. La búsqueda de soluciones innovadoras y efectivas para evaluar la usabilidad en la RX se convierte en un desafío apasionante para investigadores y profesionales del campo.

Estos hallazgos resaltan la diversidad de enfoques utilizados, destacando la relevancia de la entrevista, la evaluación heurística y la encuesta. Estas técnicas brindan diferentes enfoques para recopilar información, identificar problemas y obtener opiniones de los usuarios, lo cual es fundamental para mejorar la usabilidad y la experiencia del usuario en entornos de RX. Además, se destaca la importancia de adaptar y combinar técnicas para obtener una comprensión más completa de la experiencia del usuario en estos entornos. Al explorar y utilizar combinaciones de técnicas en la evaluación de la usabilidad en entornos virtuales, se pueden obtener evaluaciones más completas y precisas, capturando diversos aspectos de la experiencia del usuario. Además, estas combinaciones permiten abordar los desafíos y problemas específicos asociados con la evaluación de la usabilidad en este tipo de entornos. Mediante la combinación adecuada de técnicas, se puede obtener una visión integral y enriquecedora de la usabilidad, lo cual resulta fundamental para mejorar la experiencia de los usuarios en los entornos virtuales.

Para finalizar, cabe destacar que la discusión de los resultados obtenidos en esta investigación permite examinar los entornos de RX más utilizados, como la RV, y aquellos menos utilizados, como la RA y la RM. Se exploran las técnicas de evaluación de usabilidad más frecuentemente empleadas en estos entornos, analizando cómo se adaptan, combinan y crean *frameworks* ajustados a las particularidades de cada entorno. Se abordan los desafíos y retos que enfrentan los investigadores al aplicar ciertas técnicas de evaluación de usabilidad en este contexto. Estos hallazgos contribuyen a comprender el estado actual de la evaluación de usabilidad en la RX y ofrecen perspectivas valiosas para futuras investigaciones en este ámbito.

5.2. Amenazas a la Validez

En nuestro EMS identificamos tres amenazas a la validez. En primer lugar, la validez del EMS se ve amenazada por incluir únicamente estudios escritos en inglés. En segundo lugar, aunque los términos utilizados (obtenidos a partir de un GC) en la cadena de búsqueda fueron los más comúnmente usados en otros estudios del área de investigación, es posible que hayamos pasado por alto otros términos utilizados que describan trabajos relevantes.

En tercer lugar, los autores de un EMS pueden cometer errores de apreciación al seleccionar los estudios. Para contrarrestar esta amenaza, definimos criterios de inclusión y exclusión relacionados directamente con el objetivo y las preguntas propuestas. Además, para verificar la concordancia en la selección, se realizaron reuniones para comprobar los estudios preseleccionados descartados.

Otro aspecto relacionado con el proceso de selección de estudios es el alcance declarado en el presente trabajo de titulación, ya que solo consideramos trabajos publicados entre 2018 y 2022. Es probable que hayamos descartado algunos estudios relacionados directamente con el objetivo de nuestra investigación al considerar solo este periodo. Además, únicamente consideramos las bases de datos científicas Scopus, IEEE Xplore y WoS. A pesar de haber encontrado muchos resultados, se podrían haber reportado más evaluaciones de usabilidad, en otras bases de datos.

Capítulo 6

Conclusiones

El presente trabajo de titulación describe el EMS realizado para identificar el estado del arte de las técnicas utilizadas en la evaluación de la usabilidad de entornos virtuales durante el período de 2018 a 2022. Durante este período, se encontraron un total de 30 estudios primarios, mayoritariamente publicados en revistas y conferencias. Se ha observado un creciente interés en la evaluación de la usabilidad de entornos virtuales a partir de 2020. A continuación, se presentarán las conclusiones obtenidas con base en los objetivos y a las preguntas planteadas al inicio del trabajo.

RQ1: ¿Cuáles son las técnicas de usabilidad que se están utilizando para evaluar entornos virtuales? Con base en los resultados obtenidos del estudio sobre las técnicas de evaluación de usabilidad en entornos de RX, se pudo obtener una visión general de las técnicas más utilizadas en el desarrollo de entornos virtuales. Entre las técnicas más comunes para evaluar la usabilidad de estos sistemas se encuentran las entrevistas, la evaluación heurística y las encuestas (como SUS - y encuestas adaptadas). Además, se identificaron propuestas de *frameworks* que reflejan la necesidad de complementar las técnicas tradicionales para lograr una evaluación más completa. Dentro de las técnicas de evaluación de usabilidad en entornos de RX, se destaca el uso de entrevistas como una forma ampliamente utilizada para recopilar información sobre la usabilidad de estos sistemas. Esta técnica permite obtener perspectivas directas de los usuarios, comprender sus experiencias y necesidades en relación con la usabilidad de los sistemas. Las entrevistas ofrecen una valiosa perspectiva de los usuarios finales, permitiendo una comprensión más completa de la usabilidad, la satisfacción y las experiencias subjetivas. Los estudios que emplean entrevistas siguen procedimientos específicos, obtienen el consentimiento informado de los participan-

tes y utilizan escalas, cuestionarios y otros métodos para evaluar la usabilidad y la satisfacción de los usuarios. La recopilación adicional de información a través de cuestionarios y entrevistas semiestructuradas ayuda a profundizar en las percepciones y sentimientos de los participantes. La evaluación heurística también se destaca como una técnica efectiva y fácil de aprender para identificar problemas de usabilidad en entornos virtuales y aumentados. Esta técnica requiere la participación de expertos en usabilidad y proporciona un enfoque sistemático para ofrecer recomendaciones de mejora. La aplicación de heurísticas y pautas de usabilidad específicas en diferentes dominios brinda un enfoque estructurado y efectivo para mejorar la experiencia del usuario. A través de sus fases y pasos, el método heurístico establece una base sólida para identificar y abordar problemas de usabilidad, promoviendo interfaces de usuario más intuitivas y satisfactorias. Las encuestas, como el SUS y las encuestas adaptadas, también son ampliamente utilizadas para evaluar la usabilidad en entornos virtuales. Estas encuestas proporcionan datos cuantitativos sobre la satisfacción de los usuarios y se han convertido en herramientas fundamentales en el ámbito de la evaluación de usabilidad. Permiten medir y evaluar diferentes aspectos de la usabilidad, brindando información valiosa para futuros desarrollos y mejoras. Estas técnicas proporcionan una comprensión profunda de la usabilidad, la satisfacción y la experiencia del usuario en estos entornos, y permiten identificar áreas de mejora para lograr interfaces más efectivas y satisfactorias.

RQ2: ¿Cuáles son los entornos de RX que se están evaluando actualmente? Identificamos tres tipos de entornos de RX: RV, RA y MR. Al examinar los estudios primarios incluidos en esta investigación, se encontró que la mayoría de ellos (cerca del 80%) se enfocaron en la evaluación de usabilidad de sistemas software de RV. Esto indica un enfoque significativo en la evaluación de la usabilidad en la RV en comparación con la RA y la MR. En total, se identificaron 25 estudios primarios que evaluaron la usabilidad de sistemas de RV, lo cual demuestra una atención considerable en esta tecnología. Sin embargo, solo tres estudios primarios evaluaron la usabilidad de sistemas de RA, y únicamente dos estudios primarios evaluaron la usabilidad de sistemas de MR. Estos hallazgos sugieren un menor número de investigaciones centradas en estas áreas específicas de la RX.

RQ3: ¿Cuáles son los principales problemas y desafíos en la aplicación técnicas para evaluar la usabilidad de los entornos virtuales? Se identificaron en el EMS cuatro estudios primarios que reportan explícitamente retos o el éxito en la aplicación

de las técnicas de evaluación de usabilidad. Los problemas identificados se relacionan con el tamaño de la muestra, el tipo de muestreo realizado y la aplicación de las técnicas de evaluación. Específicamente, las técnicas de métricas de usabilidad y métodos híbridos presentaron dificultades en su implementación. Además, estos cuatro estudios primarios son relevantes, ya que reportaron los resultados de aplicar técnicas de usabilidad en entornos virtuales, proporcionando información valiosa sobre la efectividad de las técnicas utilizadas y los posibles problemas o desafíos encontrados durante la evaluación. Estos informes contribuyen significativamente a comprender la aplicación práctica de las técnicas de evaluación de usabilidad y su impacto en la mejora de la experiencia del usuario en los entornos virtuales. En cuanto a los siete estudios que reportan problemas de usabilidad, se han identificado una serie de desafíos significativos en la usabilidad de los entornos virtuales. Estos problemas abarcan desde cuestiones técnicas, como la retroalimentación lenta y la distorsión de la percepción, hasta obstáculos relacionados con la interfaz de usuario, como la falta de interacción funcional y la sobrecarga cognitiva. Asimismo, se han resaltado problemas de diseño que impactan la experiencia del usuario, como las dificultades en la navegación y la falta de claridad en las instrucciones. Estos hallazgos subrayan la importancia de abordar de manera minuciosa la usabilidad en el desarrollo de entornos virtuales, enfocándose en simplificar la interfaz, mejorar la retroalimentación y considerar la experiencia del usuario en su totalidad para asegurar una interacción efectiva y satisfactoria en estos entornos inmersivos.

RQ4: ¿Cómo se aplican las técnicas de usabilidad para evaluar los entornos virtuales? Durante el análisis de resultados, se identificaron cinco enfoques diferentes para evaluar la usabilidad de los entornos de RX. Estos enfoques incluyen la aplicación de técnicas de evaluación de usabilidad tradicionales según la teoría, la adaptación de estas técnicas, la combinación de diferentes técnicas de evaluación de usabilidad, la propuesta y aplicación de *frameworks*, y la propuesta y aplicación de métodos específicos de evaluación de usabilidad. En cuanto a las adaptaciones, se observó la importancia de adaptar las encuestas para medir aspectos relevantes de la usabilidad en contextos específicos. En dos estudios mencionados anteriormente, se realizaron modificaciones a las preguntas basadas en pruebas de usabilidad anteriores y marcos teóricos existentes. Esto demuestra cómo las encuestas pueden adaptarse para obtener información precisa y relevante sobre la usabilidad en entornos de RX. Otro ejemplo de adaptación es el método de recorrido cognitivo adaptado, que

implica un análisis detallado de las tareas necesarias y la generación de preguntas para cada paso. Este enfoque muestra cómo se puede adaptar un método tradicional para evaluar la usabilidad en entornos de RV. Además de las adaptaciones, se identificaron diversas combinaciones de técnicas utilizadas en conjunto para obtener una evaluación más completa de la usabilidad. Estas combinaciones incluyen la entrevista junto con la retroalimentación del usuario, la evaluación heurística, la encuesta, el pensamiento en voz alta, el recorrido cognitivo y el seguimiento ocular. Al combinar estas técnicas, se logra un panorama más completo y detallado de la experiencia del usuario, permitiendo así abordar de manera efectiva los desafíos y problemas de usabilidad que puedan surgir. Asimismo, los autores de los estudios primarios identificados en el EMS proponen y aplican *frameworks* específicos para evaluar la usabilidad en entornos de RV. Estos *frameworks* abordan diferentes aspectos de la usabilidad, como la presencia del sistema, el compromiso del público, el recuerdo de la información y el deseo de ver el contenido en directo. Estas estructuras proporcionan pautas y métricas que permiten evaluar y mejorar la usabilidad en estos entornos. Además de los *frameworks*, se identificaron propuestas y aplicación de técnicas nuevas para medir la usabilidad y la experiencia del usuario en entornos de RV. Estas metodologías permiten identificar problemas de usabilidad, analizar el comportamiento del usuario y evaluar el valor educativo de los entornos virtuales de aprendizaje. Al emplear enfoques y técnicas adaptadas a la RV, se facilita la mejora continua de las aplicaciones y entornos de RV, brindando experiencias más satisfactorias y efectivas para los usuarios. Es decir, en los entornos de RX, las adaptaciones, combinaciones y propuestas de métodos y *frameworks* son esenciales para adaptar los sistemas a diferentes dispositivos, enriquecer las experiencias de RX y facilitar el desarrollo de aplicaciones. Estas prácticas impulsan el crecimiento y la innovación en la industria de la RX, permitiendo la creación de experiencias más inmersivas y atractivas para los usuarios.

Las conclusiones del trabajo de titulación demuestran de manera clara que se han alcanzado todos los objetivos específicos que se propusieron al inicio de este proyecto. A medida que se contempla el horizonte de trabajos futuros, se reconoce la importancia de expandir la búsqueda a través de la inclusión de más bases de datos (por ejemplo, SpringerLink y ScienceDirect) con el objetivo de obtener información más completa y enriquecer aún más las respuestas a las preguntas planteadas, ya sea profundizando en el tema o contribuyendo de manera más significativa. Además,

llevar a cabo un análisis detallado para determinar si las técnicas se han aplicado conforme a las pautas establecidas en el campo de la IPC, o si por el contrario han sido adaptadas. En el caso de encontrar técnicas de evaluación de usabilidad adaptadas, se identificarán cuáles fueron las adaptaciones realizadas y qué razones motivaron tal adaptación. Posteriormente, se presentará un marco de trabajo que facilite la adaptación sistemática de estas técnicas, seguido de la evaluación de dicho marco propuesto.

Referencias

- Alpala, L. O., Quiroga-Parra, D. J., Torres, J. C., y Peluffo-Ordóñez, D. H. (2022). Smart factory using virtual reality and online multi-user: Towards a metaverse for experimental frameworks. *Applied Sciences*, 12(12):6258. 33, 34, 85
- Azuma, R., Baillet, Y., Behringer, R., Feiner, S., Julier, S., y MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47. 12
- Azuma, R. T. (1997). A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385. 35
- Barricelli, Barbara Rita and De Bonis, Ambra and Di Gaetano, Serena and Valtolina, Stefano and others (2018). Semiotic framework for virtual reality usability and ux evaluation: a pilot study. En *GHIItaly@ AVI*. 3, 34, 40, 41, 82
- Besoain, F., Jego, L., y Arenas-Salinas, M. (2018). Implementation of a gamified puzzle based on pro-origami protein structure cartoons: An experience in virtual reality. En *2018 IEEE Biennial Congress of Argentina (ARGENCON)*, pp. 1–7. IEEE. 16
- Besoain, F., Jego, L., y Gallardo, I. (2021). Developing a virtual museum: Experience from the design and creation process. *Information*, 12(6):244. 31, 33, 34, 84
- Birnie, K. A., Kulandaivelu, Y., Jibb, L., Hroch, P., Positano, K., Robertson, S., Campbell, F., Abla, O., y Stinson, J. (2018). Usability testing of an interactive virtual reality distraction intervention to reduce procedural pain in children and adolescents with cancer. *Journal of Pediatric Oncology Nursing*, 35(6):406–416. 3, 27, 28, 30, 34, 40, 41, 83

- Birt, J. y Vasilevski, N. (2021). Comparison of single and multiuser immersive mobile virtual reality usability in construction education. *Educational Technology & Society*, 24(2):93–106. 3, 29, 34, 38, 40, 41, 84
- Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7. 29
- Castro, J. W., Garnica, I., y Rojas, L. A. (2022). Automated tools for usability evaluation: A systematic mapping study. En *Social Computing and Social Media: Design, User Experience and Impact: 14th International Conference, SCSM 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I*, pp. 28–46. Springer. 1, 6, 20
- Cheng, K.-H. (2022). Teachers’ perceptions of exploiting immersive virtual field trips for learning in primary education. *Journal of Research on Technology in Education*, 54(3):438–455. 27, 28, 34, 37, 85
- Çöltekin, Arzu and Lochhead, Ian and Madden, Marguerite and Christophe, Sidonie and Devaux, Alexandre and Pettit, Christopher and Lock, Oliver and Shukla, Shashwat and Herman, Lukáš and Stachoň, Zdeněk and others (2020). Extended reality in spatial sciences: A review of research challenges and future directions. *ISPRS International Journal of Geo-Information*, 9(7):439. 35
- Constantine, L. L. y Lockwood, L. A. (1999). *Software for use: a practical guide to the models and methods of usage-centered design*. Pearson Education. 30, 31, 33
- Cooper, A., Reimann, R., y Cronin, D. (2007). *About face 3: the essentials of interaction design*. John Wiley & Sons. 32
- Dantes, G. R., Suputra, P. H., Sudarma, I. K., Suwastini, N. K. A., y Dantes, K. R. (2020). Evaluating and redesigning virtual reality’underwater tourism’application based on heuristic method. *International Journal of Business Information Systems*, 35(2):225–238. 28, 34, 35, 37, 83
- De Melo, M. S. P., da Silva Neto, J. G., Da Silva, P. J. L., Teixeira, J. M. X. N., y Teichrieb, V. (2019). Analysis and comparison of robotics 3d simulators. En *2019 21st Symposium on Virtual and Augmented Reality (SVR)*, pp. 242–251. IEEE. 3, 29, 30, 34, 40, 41, 83

- Delamarre, A., Lisetti, C., y Buche, C. (2020). A cross-platform classroom training simulator: Interaction design and evaluation. En *2020 International Conference on Cyberworlds (CW)*, pp. 86–93. IEEE. 2, 32, 34, 35, 36, 37, 83
- Doumanis, I. y Economou, D. (2019). Using cognitive walkthrough and hybrid prototyping to gather user requirements in early design virtual reality prototypes. En *Immersive Learning Research Network: 5th International Conference, iLRN 2019, London, UK, June 23–27, 2019, Proceedings 5*, pp. 234–246. Springer. 3, 28, 30, 34, 36, 38, 39, 40, 41, 83
- Fajar, M., Udjaja, Y., y David, A. C. (2022). A comparative investigation of usability issues toward virtual reality implementation in a state-owned shipping service enterprise. *ICIC Express Lett. Part B: Appl.*, 13(5):545–552. 28, 34, 84
- Ferré, X., Juristo, N., Windl, H., y Constantine, L. (2001). Usability basics for software developers. *IEEE software*, 18(1):22–29. 9, 10
- Granić, A., Nakić, J., y Marangunić, N. (2020). Scenario-based group usability testing as a mixed methods approach to the evaluation of three-dimensional virtual learning environments. *Journal of Educational Computing Research*, 58(3):616–639. 2, 4, 34, 35, 36, 37, 44, 45, 83
- Harms, P. (2019). Automated usability evaluation of virtual reality applications. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(3):1–36. 4, 34, 44, 45, 83
- Hassandra, Mary and Galanis, Evangelos and Hatzigeorgiadis, Antonis and Goudas, Marios and Mouzakidis, Christos and Karathanasi, Eleni Maria and Petridou, Niki and Tsolaki, Magda and Zikas, Paul and Evangelou, Giannis and others (2021). A virtual reality app for physical and cognitive training of older people with mild cognitive impairment: mixed methods feasibility study. *JMIR serious games*, 9(1):e24170. 27, 28, 29, 34, 84
- He, L., Li, H., Xue, T., Sun, D., Zhu, S., y Ding, G. (2018). Am i in the theater? usability study of live performance based virtual reality. En *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11. 3, 34, 40, 82

- Heiyanthuduwa, T. A., Amarapala, K. N. U., Gunathilaka, K. V. B., Ravindu, K. S., Wickramarathne, J., y Kasthurirathna, D. (2020). Virtualpt: Virtual reality based home care physiotherapy rehabilitation for elderly. En *2020 2nd International Conference on Advancements in Computing (ICAC)*, volumen 1, pp. 311–316. IEEE. 1
- Hertzum, M. (2020). Usability testing: A practitioner’s guide to evaluating the user experience. *Synthesis Lectures on Human-Centered Informatics*, 13(1):i–105. 1, 9, 10
- Hickman, L. y Akdere, M. (2017). Exploring virtual reality for developing soft-skills in stem education. En *2017 7th World Engineering Education Forum (WEEF)*, pp. 461–465. IEEE. 1, 11
- Hix, D. y Hartson, H. R. (1993). *Developing user interfaces: ensuring usability through product & process*. John Wiley & Sons, Inc. 31
- Jangid, V. y Kongsilp, S. (2021). Fishtank sandbox: A software framework for collaborative usability testing of fish tank virtual reality interaction techniques. En *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–3. 3, 34, 40, 43, 84
- Juristo, N., Windl, H., y Constantine, L. (2001). Introducing usability. *IEEE software*, 18(1):20. 10
- Kabassi, K., Maravelakis, E., y Konstantaras, A. (2018). Heuristics and fuzzy multi-criteria decision making for evaluating museum virtual tours. *International Journal of the Inclusive Museum*, 11(3). 2, 28, 34, 35, 36, 82
- Kamińska, D., Zwoliński, G., y Laska-Leśniewicz, A. (2022). Usability testing of virtual reality applications—the pilot study. *Sensors*, 22(4):1342. 3, 28, 31, 32, 34, 40, 41, 84
- Kardong-Edgren, S. S., Farra, S. L., Alinier, G., y Young, H. M. (2019). A call to unify definitions of virtual reality. *Clinical Simulation in Nursing*, 31:28–34. 1, 11, 12, 13

- Karre, S. A., Mathur, N., y Reddy, Y. R. (2020). Understanding usability evaluation setup for vr products in industry: A review study. *ACM SIGAPP Applied Computing Review*, 19(4):17–27. 13, 14
- Kitchenham, B. A., Budgen, D., y Brereton, O. P. (2011). Using mapping studies as the basis for further research—a participant-observer case study. *Information and Software Technology*, 53(6):638–651. 6, 15
- Li, C., Liang, W., Quigley, C., Zhao, Y., y Yu, L.-F. (2017). Earthquake safety training through virtual drills. *IEEE transactions on visualization and computer graphics*, 23(4):1275–1284. 1
- Li, Z., Wang, J., Yan, Z., Wang, X., y Anwar, M. S. (2019). An interactive virtual training system for assembly and disassembly based on precedence constraints. En *Advances in Computer Graphics: 36th Computer Graphics International Conference, CGI 2019, Calgary, AB, Canada, June 17–20, 2019, Proceedings 36*, pp. 81–93. Springer. 28, 29, 34, 83
- Liu, Y., Yang, N., Li, A., Paterson, J., McPherson, D., Cheng, T., y Yang, A. Y. (2018). Usability evaluation for drone mission planning in virtual reality. En *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part II 10*, pp. 313–330. Springer. 3, 34, 36, 40, 41, 42, 82
- Losana, P., Castro, J. W., Ferre, X., Villalba-Mora, E., y Acuña, S. T. (2021). A systematic mapping study on integration proposals of the personas technique in agile methodologies. *Sensors*, 21(18):6298. 1, 6
- Magües, D. A., Castro, J. W., y Acuña, S. T. (2016). Usability in agile development: A systematic mapping study. En *2016 XLII Latin American Computing Conference (CLEI)*, pp. 1–8. IEEE. 20
- Mayhew, D. J. (1999). The usability engineering lifecycle. En *CHI'99 Extended Abstracts on Human Factors in Computing Systems*, pp. 147–148. 30, 32
- Milgram, P. y Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329. 35

- Mitre-Ortiz, A., Muñoz-Arteaga, J., y Cardona-Reyes, H. (2022). Developing a model to evaluate and improve user experience with hand motions in virtual reality environments. *Universal Access in the Information Society*, pp. 1–15. 28, 34, 85
- Moore, A. G., Hu, X., Eubanks, J. C., Aiyaz, A. A., y McMahan, R. P. (2020). A formative evaluation methodology for vr training simulations. En *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 125–132. IEEE.
- Naranjo, J. E., Urrutia, F. U., Garcia, M. V., Gallardo-Cárdenas, F., Franklin, T. O., y Lozada-Martínez, E. (2019). User experience evaluation of an interactive virtual reality-based system for upper limb rehabilitation. En *2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG)*, pp. 328–333. IEEE. 16
- Nielsen, J. (1994a). Heuristic evaluation. *Usability Inspection Methods*. 28
- Nielsen, J. (1994b). *Usability engineering*. Morgan Kaufmann. 27, 30, 31
- Nielsen, J. y Molich, R. (1990). Heuristic evaluation of user interfaces. En *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 249–256. 32
- O’Grady, J. V. y O’Grady, K. V. (2017). *A Designer’s Research Manual, Updated and Expanded: Succeed in Design by Knowing Your Clients and Understanding what They Really Need*. Rockport Publishers. 33
- Ogunseiju, O. R., Gonsalves, N., Akanmu, A. A., Bairaktarova, D., Bowman, D. A., y Jazizadeh, F. (2022). Mixed reality environment for learning sensing technology applications in construction: A usability study. *Advanced Engineering Informatics*, 53:101637. 3, 30, 31, 32, 33, 34, 35, 40, 41, 85
- Oliveira, E., Simões, F. P., y Correia, W. F. (2017). Heuristics evaluation and improvements for low-cost virtual reality. En *2017 19th Symposium on Virtual and Augmented Reality (SVR)*, pp. 178–187. IEEE. 16
- Özdingç, F., Tüzün, H., Ergün, E., Bayrak, F., y Kula, A. (2016). Usability testing of a three-dimensional library orientation game. *Games User Research: A Case Study Approach*, pp. 77–95. 16

- Pedroli, Elisa and Greci, Luca and Colombo, Desirée and Serino, Silvia and Cipresso, Pietro and Arlati, Sara and Mondellini, Marta and Boilini, Lorenzo and Giussani, Valentina and Goulene, Karine and others (2018). Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: Positive bike. *Sensors*, 18(7):2343. 16, 28, 29, 34, 82
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., y Carey, T. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd. 27, 28, 29, 30, 31, 33
- Ren, R., Castro, J. W., Acuña, S. T., y de Lara, J. (2019). Evaluation techniques for chatbot usability: A systematic mapping study. *International Journal of Software Engineering and Knowledge Engineering*, 29(11n12):1673–1702. 20
- Ren, R., Zapata, M., Castro, J. W., Dieste, O., y Acuña, S. T. (2022). Experimentation for chatbot usability evaluation: A secondary study. *IEEE Access*, 10:12430–12464. 1
- Rohrbach, N., Gulde, P., Armstrong, A. R., Hartig, L., Abdelrazeq, A., Schröder, S., Neuse, J., Grimmer, T., Diehl-Schmid, J., y Hermsdörfer, J. (2019). An augmented reality approach for adl support in alzheimer’s disease: a crossover trial. *Journal of neuroengineering and rehabilitation*, 16:1–11. 3, 28, 29, 34, 35, 40, 41, 83
- Rzeszewski, M. y Orylski, M. (2021). Usability of webxr visualizations in urban planning. *ISPRS International Journal of Geo-Information*, 10(11):721. 11, 28, 29, 34, 35, 37, 84
- Schmidt, M., Schmidt, C., Glaser, N., Beck, D., Lim, M., y Palmer, H. (2021). Evaluation of a spherical video-based virtual reality intervention designed to teach adaptive skills for adults with autism: A preliminary report. *Interactive Learning Environments*, 29(3):345–364. 3, 28, 29, 30, 34, 40, 41, 84
- Shackel, B. (2009). Usability–context, framework, definition, design and evaluation. *Interacting with computers*, 21(5-6):339–346. 9
- Sheikh, S., Heyat, M. B. B., AlShorman, O., Masadeh, M., y Alkhatni, F. (2021). A review of usability evaluation techniques for augmented reality systems in education. En *2021 Innovation and New Trends in Engineering, Science and Technology Education Conference (IETSEC)*, pp. 1–6. IEEE. 13, 14

- Shneiderman, B. (1998). Designing the user interface. 28, 29, 30
- Strada, F., Bottino, A., Lamberti, F., Mormando, G., e Ingrassia, P. L. (2019). Holo-blsc—a holographic tool for self-training and self-evaluation of emergency response skills. *IEEE Transactions on Emerging Topics in Computing*, 9(3):1581–1595. 29, 34, 35, 84
- Thalmann, M., Ringli, L., Adcock, M., Swinnen, N., De Jong, J., Dumoulin, C., Guimarães, V., y de Bruin, E. D. (2021). Usability study of a multicomponent exergame training for older adults with mobility limitations. *International journal of environmental research and public health*, 18(24):13422. 4, 35, 40, 44, 84
- Tham, J., Duin, A. H., Gee, L., Ernst, N., Abdelqader, B., y McGrath, M. (2018). Understanding virtual reality: Presence, embodiment, and professional practice. *IEEE Transactions on Professional Communication*, 61(2):178–195. 1, 11
- Tromp, J., Wolff, A., Torres, J., y My, H. (2018). Usability evaluation of the interactive 3d virtual cultural heritage museum display: Fountain of the lions software application. *International Journal of Engineering and Technology(UAE)*, 7:95–99. 3, 28, 30, 31, 32, 34, 35, 40, 41, 82
- Verkuyl, M., Romaniuk, D., y Mastrilli, P. (2018). Virtual gaming simulation of a mental health assessment: A usability study. *Nurse Education in Practice*, 31:83–87. 2, 3, 28, 29, 31, 34, 35, 36, 38, 40, 41, 83
- Zanatta, F., Giardini, A., Pierobon, A., D’Addario, M., y Steca, P. (2022). A systematic review on the usability of robotic and virtual reality devices in neuromotor rehabilitation: patients’ and healthcare professionals’ perspective. *BMC health services research*, 22(1):523. 13, 14
- Zhang, H., Babar, M. A., y Tell, P. (2011). Identifying relevant studies in software engineering. *Information and Software Technology*, 53(6):625–637. 16

Anexo A - Publicación Derivada

En este anexo, se presenta el artículo derivado del trabajo de titulación, el que fue aceptado para su presentación en la *25th International Conference on Human-Computer Interaction*, que fue realizada de forma remota el 23 de julio del presente año.



Usability Evaluation Techniques for Virtual Environments: An Exploratory Study

John W. Castro¹(✉), Gianina Madrigal¹, and Luis A. Rojas²

¹ Departamento de Ingeniería Informática Y Ciencias de La Computación, Universidad de Atacama, Copiapó, Chile

john.castro@uda.cl, gianina.madrigal.17@alumnos.uda.cl

² Departamento de Ciencias de La Computación Y Tecnologías de La Información, Universidad del Bío-Bío, Chillan, Chile

lurojas@ubiobio.cl

Abstract. Virtual reality is a concept of specialized hardware that generates an experience with different levels of intensity and interactivity within environments in which users feel immersed in an apparently real world and is defined based on the technology. Currently, several fields, such as education and health, use virtual reality. Due to the wide range of users (children, youth, adults, and even older adults) who can interact with virtual environments, it is essential to consider the usability of this type of software system. The evaluation of usability is carried out through usability tests. Research on evaluating virtual environments' usability is scattered in several scientific databases. To the best of our knowledge, no research work compiles and reports globally on which techniques are used to evaluate the usability of such software systems. We conducted a systematic mapping study to assess the general panorama of techniques related to usability evaluation that are being used in virtual environments. We identified a total of 30 primary studies. The main techniques for usability evaluation are interviews, heuristic evaluation, surveys, usability experiments and expert evaluation. Some of these techniques are being used together. In addition, we have identified some works that propose frameworks or methodologies for evaluating the usability of virtual reality systems.

Keywords: Usability Evaluation · Virtual Environments · Exploratory Study

1 Introduction

Virtual reality (VR) is a concept of specialized hardware that generates an experience with different levels of intensity and interactivity within environments in which users feel immersed in an apparently real world and is defined based on the technology [1, 2]. Currently, several fields, such as education [3], health [4], and seismic safety [5], use virtual reality. Due to the wide range of users (children, youth, adults, and even older adults) who can interact with virtual environments, it is essential to consider the usability of this type of software system, which increases efficiency and user satisfaction, also improving their productivity [6].

Usability is one of the most critical indicators in determining the quality of a software product [7–9]. It corresponds to how a group of users can use a software system to achieve specific objectives with effectiveness, efficiency, and satisfaction [6]. The evaluation of usability is carried out through usability tests since, despite the efforts of software developers, the first designs tend to be flawed since they can, for example, [6]: (i) Assume knowledge that users do not possess, (ii) use confusing or highly technical vocabulary, and (iii) impose excessive mental workload on users. To avoid these situations, it is necessary to evaluate usability so that the user will not feel dissatisfied with the final product [6]. Research on evaluating virtual environments' usability is scattered in several scientific databases.

To the best of our knowledge, no research work compiles and reports globally on which techniques are used to evaluate the usability of such software systems. Therefore, professionals and researchers who need this information will have to: (i) Spend much time searching and finding relevant sources of knowledge, (ii) review the resources obtained according to specific criteria, (iii) extract information from multiple sources, and, finally, (iv) conclude from the information found. This problem becomes more acute when it is considered that the scientific community periodically contributes new bibliographic resources.

This research seeks to assess the general panorama of techniques related to usability evaluation that are being used in virtual environments. For this, we conducted a systematic mapping study (SMS). The databases used are Scopus, IEEE Xplore, and Web of Science (WoS). We identified a total of 30 primary studies. The main techniques for usability evaluation are interviews, heuristic evaluation, surveys, usability experiments and expert evaluation. Some of these techniques are being used together. In addition, we have identified some works that propose frameworks or methodologies for evaluating the usability of virtual reality systems [10, 11].

This paper is organized as follows. In Sect. 2, we present the related work. Section 3 describes the research method (i.e., SMS). In Sect. 4, we discuss the results of the SMS. Section 5 presents possible threats to validity, and finally, the conclusions are presented in Sect. 6.

2 Related Work

As a result of our pilot search, we found three studies [12–14] related to our research. Karre et al. [12] report usability studies carried out in the industry when developing VR products. Karre et al. [12] focused on those studies that design VR scenes precisely and develop a simple protocol to perform or set up a setup for the usability evaluation of those scenes. This study aims to contribute when choosing a usability evaluation method for future VR products at an industrial level. Although this study is complete, it is outdated since it was carried out from 2000 to 2018.

The second work [13] reports specific usability evaluation techniques in augmented reality (AR) systems, specifically in education. Sheikh et al. [13] state that some critical challenges must be considered to implement RA, especially in the educational area. In this area, the users may require a greater demand for motor skills due to the interaction with 3D projections. In this work, the authors report a miscellany of usability techniques

that are pretty efficient when evaluating AR systems for education. While this could be extended to other areas that use such technology, it does not focus on all applicable areas of AR technology.

In the latest study, Zanatta et al. [14] report works that evaluate the usability of VR systems and robotic devices applied to neuromotor rehabilitation. The authors focus on a variety of essential areas to be able to implement a VR system in the area of rehabilitation, among which usability is mentioned. In addition, this research addresses the usability techniques that were visualized in the primary studies identified in their research. Although the study by Zanatta et al. [14] provides updated and complete information and provides some recommendations, it is not generic enough since it focuses on one of the many areas in which VR is applied.

After carrying out an analysis of the studies mentioned above, it is possible to point out that the present research paper differs from the last two (i.e., [13] and [14]). On the one hand, this research is not limited to particular areas, and not only VR is considered, but also AR and mixed reality (MR) to obtain a broader vision of how the usability of systems is being evaluated that use these technologies. On the other hand, our work differs from the first study described (i.e., [12]) since we carried out an updated investigation (from 2018 to September 2022), and the focus of the research questions is different. Therefore, this research work is necessary since VR, AR, and MR technologies are constantly growing, and knowing which techniques are used to evaluate their usability correctly is necessary.

3 Research Method

Software engineering considers usability as one of the key attributes of software [7, 8]. *Usability* is a measure that can be used in any product without being exclusive to computer systems. The secondary study reported in this work has been developed following the guidelines established by Kitchenham et al. [15] for carrying out an SMS. Following these guidelines, the following activities were carried out: (i) formulation of the research questions; (ii) definition of the search strategy; (iii) selection of primary studies, (iv) extraction; (v) and synthesis of the extracted data.

3.1 Research Questions

The information extracted from the primary studies aims to answer the following research questions (RQ): (RQ1) What are the usability techniques used to evaluate virtual environments? (RQ2) What are the main problems and challenges in applying techniques to evaluate the usability of virtual environments? (RQ3) What are the extended reality environments currently being evaluated?

3.2 Define the Search Strategy

The SMS begins with identifying the keywords, so it is necessary to identify a Control Group (CG). The CG is defined as a set of studies specifically related to the area of interest. In addition, the studies belonging to the CG must represent the research as

accurately as possible and answer the research questions formulated [16]. The CG arises from the need for objectivity in selecting the search string.

A manual search of studies related to the research context is performed to form the CG. Likewise, these studies should answer the research questions. As a result of the search, five primary studies were identified [17–21]. Before building the search string, it is verified that the CG studies are in the Scopus database since it has the most significant number of studies. As a result of this verification, all the CG studies are found in Scopus, which is why it is the best option for research.

To obtain the keywords that will make up the search string, the Atlas.ti program (version 22) is used. Based on the above, a table was generated with the frequency of all the words and combinations that appeared in the CG articles. Only those words that were directly related to the research questions and that were present in a significant percentage of the articles belonging to the CG were chosen (we call this percentage coverage). Subsequently, each word obtained was assigned a value from 0 to 1, determined by its frequency of use, so that the most repeated word in the CG studies had the value 1. Table 1 shows a fragment of the list of words obtained due to this selection process. It shows the words, their coverage, frequency of use, and assigned weight (calculated based on coverage and frequency of use – see Eq. 1).

$$\text{Weight} = ((\text{Wordcoverage}) / (\text{Maximumcoverage}) + (\text{Wordfrequency}) / (\text{Maximumfrequency})) / 2 \quad (1)$$

Table 1. Fragment of the list of words obtained from the selection process.

Words	Coverage (%)	Frequency	Weight
Usability	100	222	1
Evaluation	100	123	0.7770
VR	100	101	0.7275
Virtual reality	100	90	0.7027
Study	100	59	0.6329
Evaluations	60	38	0.3856
Test	100	34	0.5766
Studies	80	28	0.4631
Assessment	60	19	0.3428

3.3 Formation of the Search String

After obtaining the keywords, several search strings were built, for which three components related to the context of the investigation were considered (i.e., to identify the current panorama of how the usability evaluation of virtual environments is carried out).

The defined components were the following: (i) evaluation, (ii) usability, and (iii) virtual environments. The logical operator AND was used to join each of the components and OR to include synonyms of the terms of the same component. Six strings were built to search for CG studies in the Scopus database. Table 2 shows the number of studies found, and the number of CG studies found for each string tested.

Table 2. Search strings.

ID	Search string	Studies found	GC found	Ratio X	Ratio Y	Average
1	(study OR evaluation OR studies OR assessment OR evaluations OR test) AND usability AND (VR OR “virtual reality” OR “virtual environment” OR immersive)	1955	5	1	0.00256	0.50128
2	(study OR evaluation OR assessment OR test) AND usability AND (VR OR “virtual reality” OR “virtual environment” OR immersive)	1801	5	1	0.00278	0.50139
3	(study OR evaluation OR assessment OR test) AND usability AND (VR OR “virtual reality” OR “virtual environment”)	1833	5	1	0.00273	0.50136
4	(study OR evaluation OR assessment) AND (usability) AND (VR OR “virtual reality” OR immersive)	1777	5	1	0.00281	0.50141
5	(study OR evaluation OR assessment) AND (usability) AND (VR OR “virtual reality”)	1666	4	0.8	0.00300	0.40150
6	(study OR evaluation OR studies OR assessment OR evaluations OR test) AND usability AND (VR OR “virtual reality”)	1764	5	1	0.00284	0.50142

As seen in Table 2, five of the six search strings found all five CG studies, making it necessary to search for a method to select one of the strings. This method consists of calculating the proportions X (see Eq. 2) and Y (see Eq. 3) and the average between

them (see Eq. 4).

$$X \cdot Ratio = \frac{(No.ofarticlesfoundinthecontrolgroup)}{(Totalofarticlesinthecontrolgroup)} \quad (2)$$

$$Y \cdot Ratio = \frac{(No.ofarticlesfoundfromthecontrolgroup)}{(Totalofarticlesfoundpersearchstring)} \quad (3)$$

$$Average = \frac{(XRatio + YRatio)}{2} \quad (4)$$

As observed in Table 2, the X proportion varies only once since only one of the strings did not find all the CG studies in Scopus. However, proportion Y shows specific differences since it is based on calculating the proportion of the CG studies found in the total results obtained by each string. To ensure that the selected string is the ideal one for our investigation, the mean between the X and Y ratio is calculated. Table 2 shows that string 6 has the highest mean, so it is selected as the most suitable string. Table 3 shows the structure of the final search string.

Table 3. Final search string.

Keywords				
study OR evaluation OR studies OR assessment OR evaluations OR test	AND	usability	AND	VR OR “virtual reality”

Although the tests of the strings were carried out in the Scopus database, a search was still carried out in IEEE Xplore and WoS with the final selected string to find more results. This search considered studies from 2018 to June 14, 2022. The databases were analyzed sequentially using the search fields shown in Table 4. The search fields used were determined by the options provided by each database due to the different query syntaxes [22–24]. If a duplicate appeared, the first result was kept.

Table 4. Search field per database.

Database	Search fields	Number of results
Scopus	“Title OR Abstract OR Keywords”	1425
IEEE Xplore	“Abstract”	198
Web of Science	“Title OR Abstract OR Keywords”	1035

3.4 Inclusion and Exclusion Criteria

This section presents the inclusion and exclusion criteria for selecting the primary studies. The inclusion criteria are the following:

- The study indicates the techniques used during the usability evaluation; AND

- The study reports in detail the usability evaluation or a usability study; OR
- The study reports in detail a method/technique/tool proposal to evaluate the usability of virtual reality systems.
- Regarding the exclusion criteria, they are the following:
 - The study only uses a survey or questionnaire to evaluate the usability.
 - The study does not mention the usability evaluation technique used in the abstract.
 - The study reports pilot or preliminary usability tests.
 - The study is written in a language other than English.

It should be considered that it is sufficient for a study to meet one of the exclusion criteria in order not to be considered.

3.5 Select the Studies

A total of 2658 studies were found in the three databases used. After excluding duplicate studies, the number was reduced to 1781. Subsequently, a selection of studies was made by applying the inclusion and exclusion criteria to the title and abstract of each non-duplicate study, reducing the number to 289 pre-selected. Finally, the selection criteria were strictly applied to the full text of the pre-selected studies. Figure 1 shows the entire filtering and analysis process with the inclusion and exclusion criteria used to select 30

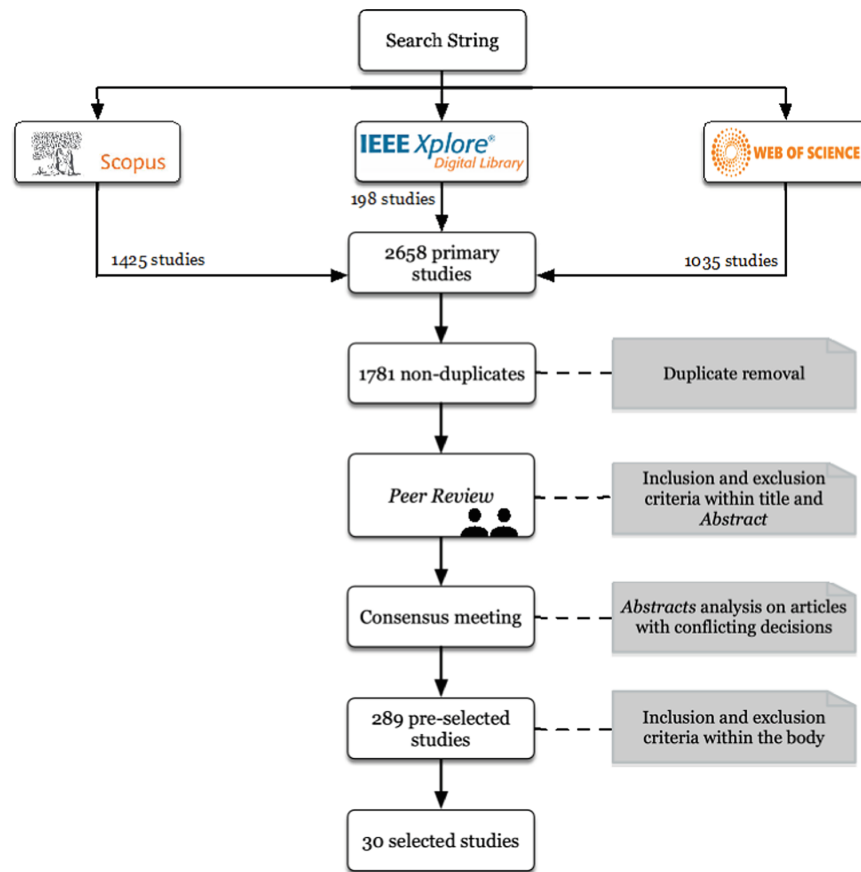


Fig. 1. Steps followed during the SMS.

primary studies. The complete list of the primary studies can be found in Annex A. The results of applying the different filters during the selection process for each database can be seen in Table 5.

Table 5. Number of remaining studies after filtering the database results.

Database	Studies found	Duplicate-free	Pre-selected studies	Primary studies
Scopus	1425	1418	244	25
IEEE Xplore	198	26	8	0
Web of Science	1035	337	37	5
Total	2658	1781	289	30

4 Results and Discussion

This section presents the results obtained in the SMS, divided into two parts: An overview of the primary studies and the answers to the research questions.

The first part is summarized in Fig. 2, which presents a synthetic view of the identified primary studies. In this figure, the results have been segmented into two areas. The first (left side) consists of two XY scatterplots (top and bottom) with bubbles at the intersections of the categories type-year of publication (left side - top) and type of publication-techniques/frameworks for evaluation usability (left side - bottom). The publication types are journals, book chapters, conferences, and workshops. The size of each bubble is determined by the number of primary studies classified as belonging to each category. The second area (right side - top) of Fig. 2 presents the number of primary studies by year of publication. As can be seen, interest in evaluating the usability of virtual environments decreased in 2019 and 2020. Subsequently, interest increased in 2021 and only decreased slightly in 2022, although for this year, studies up to September.

The second part of this section presents the synthesis of the analysis of the primary studies from which the research questions were answered.

4.1 Usability Evaluation Techniques

In this section, the research question *What are the usability techniques used to evaluate virtual environments?* is answered. From the analysis of the primary studies, 15 usability evaluation techniques were identified that are applied (see Table 6). In Table 6, the usability evaluation techniques are ordered from the most to the least used.

In addition to the techniques used for usability evaluation, some framework proposals [PS1, PS4, PS5, PS12, PS15, PS20, PS23] were identified to carry out such activity. The proposed frameworks are different in each of the primary studies. For example, the study by Harms [PS12] describes an automated VR usability evaluation method. It is important to note that this method has been applied before in desktop applications

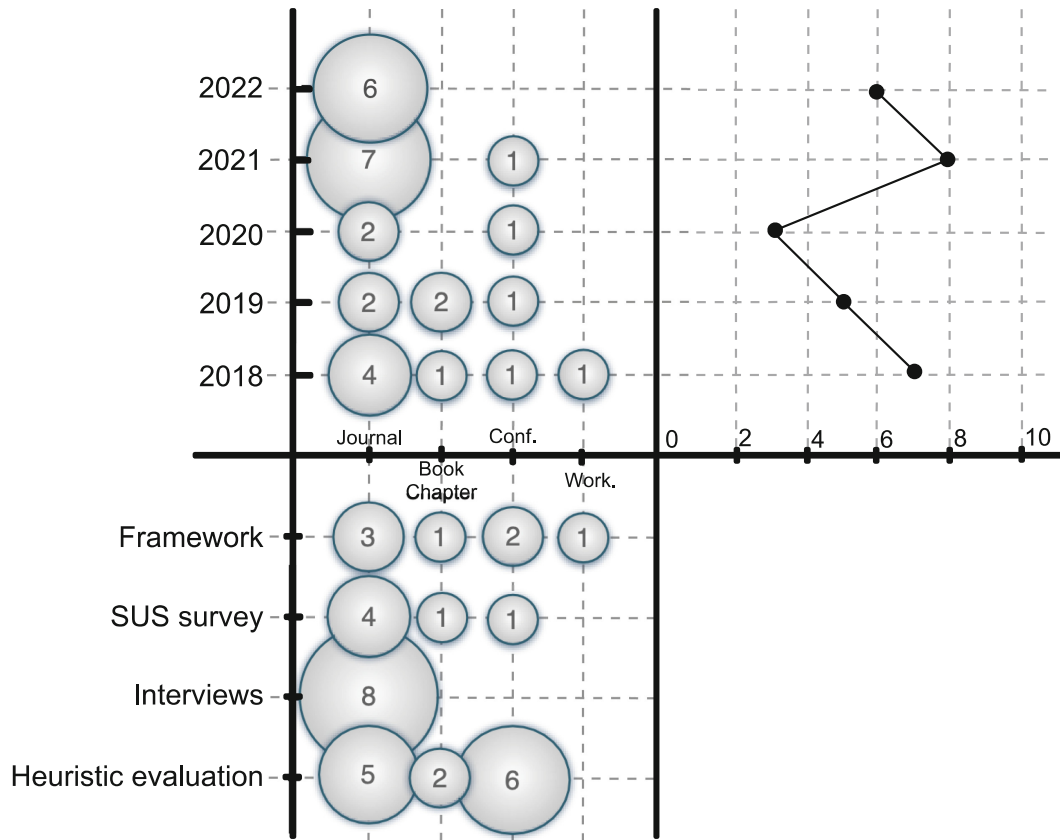


Fig. 2. Mapping for primary study distribution between usability evaluation techniques along with publication type.

and websites; that is, this method has been adapted. The approach is based on three steps. In the last step, usability smells are detected from the generated task trees. Harms [PS12] consider usability smells to be certain user behaviors that indicate an underlying usability problem. Some usability smells are Missing User Guidance, High GUI Element Distance, and Required Inefficient Actions.

4.2 Problems and Challenges in Applying Usability Evaluation Techniques

This section answers the second research question *What are the main problems and challenges in applying techniques to evaluate the usability of virtual environments?* Upon exhaustive inspection, it was found that only seven of the primary studies [PS3, PS5, PS8, PS13, PS15, PS16, PS28] report the results (i.e., whether the application of the technique was successful or, on the contrary, there were problems) or challenges) to apply the usability techniques or frameworks proposed by the authors.

It is essential to mention usability problems related to the sample size [PS5] and the type of sampling performed [PS8]. In the first case, the samples can be small. In the second case, sampling is done for convenience, a non-probabilistic or random method, where the researcher chooses the sample by proximity and without considering whether it is representative. Another problem, indicated in the study by Verkuyl et al. [PS8], is related to the number of rounds executed in the usability evaluations. However, this is

Table 6. Usability Evaluation Techniques.

Usability Techniques	Primary studies
Interviews	[PS2][PS7][PS9][PS19][PS22][PS24][PS26][PS29]
Heuristic evaluation	[PS6][PS8][PS10][PS13][PS14][PS25][PS30]
SUS survey	[PS2][PS10][PS11][PS19][PS22][PS24]
Adapted survey	[PS8][PS16][PS17][PS21]
Usability experiment	[PS9][PS10][PS17]
Expert evaluation	[PS3][PS11][PS19]
Cognitive walkthrough	[PS6][PS13]
User feedback	[PS7][PS27]
Thinking aloud	[PS8][PS27]
Usability test recording	[PS26][PS27]
Questionnaires	[PS26][PS27]
Group-based usability testing	[PS15][PS16]
Persons	[PS6]
Eye Tracking	[PS27]
Usability metrics	[PS28]

linked to the budget and time of the projects. Likewise, another challenge related to the sample is identifying the representative subjects necessary for developing VR software [PS15].

Likewise, problems and challenges are directly related to the application of usability evaluation techniques [PS3, PS13, PS16, PS26]. In the case of the expert evaluation technique [PS3], the main problem lies in the subjectivity of said evaluation since it is still an opinion, where the experts give their impressions over a well-founded opinion.

The first challenge is found when applying the technique of “group-based usability testing” [PS16]. This corresponds to the organization of the scenarios/environments and the high dropout rates of the participants. An additional limitation in the usability evaluation reported in the study is that the authors carried out a pretest to collect the participants’ demographic data, which could cause some preparation in the sample of participants, making the results unreliable.

Also, we find usability metrics as an evaluation technique. The application formula of this technique is carried out through direct or indirect questionnaires for data collection. This generates a drawback related to subjectivity since the answers tend to be related to the participants’ impressions rather than actual experiences. In addition, the questionnaires are carried out post-test, which tends to distort the answers, since when answering them after carrying out the tests with the software for a while, the experiences become blurred and are not entirely true [PS28].

Finally, the problems associated with evaluating usability through hybrid methods are related to the increase in the overall cost of usability evaluation and the complexity of applying them [PS13].

4.3 Evaluated Extended Reality Environments

This section answers the third research question *What are the extended reality environments currently being evaluated?* After conducting an analysis of the primary studies, three types of extended reality environments were identified: VR, AR and MR.

A total of 25 primary studies report the usability evaluation of VR software systems [PS1-PS8, PS10-PS13, PS15-PS20, PS23-PS26, PS28-PS30]. VR is defined as a reality in which the participants are fully immersed in a fictional world and can interact with that reality. That world can simulate the properties of specific real-world environments, whether existing or fictional. In addition, to be able to immerse yourself in this virtual reality, display devices are necessary, which function as barriers to keeping the user within this virtual world [25, 26].

Only three primary studies [PS9, PS14, PS21] report usability evaluation of AR software systems. AR is a variation of VR, and AR is a technology that unites real space and objects with computer-generated (virtual) objects, which are superimposed on the real world in real-time. The main difference between VR and AR is that AR does not fully replace the real world [26, 27].

Concerning MR, only two primary studies [PS22, PS27] report the usability evaluation of software systems with this technology. The MR unites objects of the virtual type with other objects in the real world. The main difference between MR and AR is that, in the case of the latter, virtual objects are not only superimposed on the real world but can also be interacted with. MR digital objects are tied to real space. For example, a hologram can be seen superimposed on a specific object. On the contrary, in AR, a hologram of a specific object can be placed on any flat surface [27].

5 Validity Threats

In our SMS we identified three threats to validity. First, the validity of the SMS is threatened by only including studies written in English. Second, although the terms used (derived from a CG) in the search string were the most commonly used in other studies in the research area, other terms used to describe relevant work may have been overlooked.

Third, the authors of an SMS may make judgmental errors when selecting the studies. We define inclusion and exclusion criteria directly related to the objective and research questions to counteract this threat. In addition, to verify the concordance in the selection of the studies, meetings were held between the investigators to check the discarded preselected studies.

Another aspect related to the study selection process is the declared scope of our research since only papers published between 2018 and 2022 were considered. Some studies directly related to the objective of our research have likely been discarded, considering only this period. In addition, only the scientific databases Scopus, IEEE Xplore,

and WoS were considered. Despite having found many results, more usability evaluations could have been reported.

6 Conclusions

This research paper describes the SMS carried out to identify the state of the art of techniques that are being used to evaluate the usability of virtual environments. From 2018 to 2022, we found 30 primary studies published mainly in journals and conferences. Interest in evaluating the usability of virtual environments has increased since 2020. Below we report the conclusions based on the research questions posed above.

RQ1: What are the usability techniques used to evaluate virtual environments?

According to the results of the SMS, it was possible to know the general panorama of the usability evaluation techniques used in developing virtual environments. The most commonly used techniques for evaluating the usability of this type of software system are: interviews, heuristic evaluation, and surveys (SUS survey and adapted survey). Also noteworthy are usability experiments and expert evaluation. In addition to usability evaluation techniques, we identified some framework proposals, reflecting that traditional techniques seem insufficient in some cases.

RQ2: What are the main problems and challenges in applying techniques to evaluate the usability of virtual environments?

Only seven primary studies report challenges or whether the application of the technique was successful. The problems identified correspond to the sample size, the sampling type, and the application of usability evaluation techniques. The techniques that presented problems in their application were usability metrics and hybrid methods.

RQ3: What are the extended reality environments currently being evaluated?

We identified three types of extended reality environments: VR, AR, and MR. Most of the primary studies (close to 80%) identified in the SMS report usability evaluation of VR software systems.

As future work, we consider including more databases (e.g., SpringerLink and ScienceDirect). In addition, we will analyze if the techniques have been applied as prescribed by the Human Computer-Interaction area or if, on the contrary, they have been adapted. In the case of finding adapted usability evaluation techniques, we will identify which adaptations were made and what reasons motivated such adaptations.

Acknowledgment. This work was supported in part by the ANID FONDECYT-Iniciación project under Grant 11230496 and in part by the Universidad de Atacama.

Appendix a: Primary Studies

This appendix lists the references of the primary studies used for the SMS described in this paper.

[PS1] He, L., Li, H., Xue, T., Sun, D., Zhu, S., Ding, G.: Am I in the theater? Usability study of live performance based virtual reality. In: ACM Symposium on Virtual Reality Software and Technology (VRST' 18), pp. 1–11, Tokyo, Japan (2018).

[PS2] Pedroli, E., Greci, L., Colombo, D., Serino, S., Cipresso, P., Arlati, S., Mondellini, M., Boilini, L., Giussani, V., Goulene, K., Agostoni, M., Sacco, M., Stramba-Badiale, M., Riva, G., Gaggioli, A.: Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: Positive bike. *Sensors*, 18(7), 2343 (2018).

[PS3] Kabassi, K., Maravelakis, E., Konstantaras, A.: Heuristics and fuzzy multi-criteria decision making for evaluating museum virtual tours. *International Journal of the Inclusive Museum*, 11(3), 1–21 (2018).

[PS4] Barricelli, B. R., De Bonis, A., Di Gaetano, S., Valtolina, S.: Semiotic framework for virtual reality usability and ux evaluation: A pilot study. In: 2nd Workshop on Games-Human Interaction (GHIItaly'18), pp.1–6, Castiglione della Pescaia, Grosseto, Italy (2018).

[PS5] Liu, Y., Yang, N., Li, A., Paterson, J., McPherson, D., Cheng, T., Yang, A. Y.: Usability Evaluation for Drone Mission Planning in Virtual Reality. In: Chen, J. and Fragomeni, G. (eds.). *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry. VAMR 2018. Lecture Notes in Computer Science*, vol 10910, pp. 313–330. Springer, Cham (2018).

[PS6] Tromp, J. G., Wolff, A., Torres, J. C., My, H. T.: Usability evaluation of the interactive 3D virtual cultural heritage museum display: Fountain of the lions software application. *International Journal of Engineering and Technology*, 7(2.28), 95–99 (2018).

[PS7] Birnie, K. A., Kulandaivelu, Y., Jibb, L., Hroch, P., Positano, K., Robertson, S., Campbell, F., Abla, O., Stinson, J.: Usability testing of an interactive virtual reality distraction intervention to reduce procedural pain in children and adolescents with cancer. *Journal of Pediatric Oncology Nursing*, 35(6), 406–416 (2018).

[PS8] Verkuyl, M., Romaniuk, D., Mastrilli, P.: Virtual gaming simulation of a mental health assessment: A usability study. *Nurse Education in Practice*, 31, 83–87 (2018).

[PS9] Rohrbach, N., Gulde, P., Armstrong, A. R., Hartig, L., Abdelrazeq, A., Schröder, S., Neuse, J., Grimmer, T., Diehl-Schmid, J., Hermsdörfer, J.: An augmented reality approach for ADL support in Alzheimer's disease: A crossover trial. *Journal of Neuro Engineering and Rehabilitation*, 16, 1–16, article 66 (2019).

[PS10] Li, Z., Wang, J., Yan, Z., Wang, X., Anwar, M. S.: An interactive virtual training system for assembly and disassembly based on precedence constraints. In: Gavrilova, M., Chang, J., Thalmann, N., Hitzer, E., Ishikawa, H. (eds) *Advances in Computer Graphics. CGI 2019. Lecture Notes in Computer Science*, vol 11542, pp. 81–93. Springer, Cham (2019).

[PS11] De Melo, M. S. P., Da Silva Neto, J. G., Da Silva, P. J. L., Teixeira, J. M. X. N., Teichrieb, V.: Analysis and comparison of robotics 3D simulators. In: 21st Symposium on Virtual and Augmented Reality (SVR'19), 242–251, Rio de Janeiro, Brazil (2019).

[PS12] Harms, P.: Automated usability evaluation of virtual reality applications. *ACM Transactions on Computer-Human Interaction*, 26(3), article 14 (2019).

[PS13] Doumanis, I., Economou, D.: Using cognitive walkthrough and hybrid prototyping to gather user requirements in early design virtual reality prototypes. In: Beck, D., Peña-Rios, A., Ogle, T., Economou, D., Mentzelopoulos, M., et al. (eds.) *Immersive Learning Research Network. iLRN 2019. Communications in Computer and Information Science*, vol 1044, pp. 234–246. Springer, Cham (2019).

[PS14] Dantes, G. R., Suputra, P. H., Sudarma, I. K., Suwastini, N. K. A., Dantes, K. R.: Evaluating and redesigning virtual reality ‘underwater tourism’ application based on heuristic method. *International Journal of Business Information Systems*, 35(2), 225–238 (2020).

[PS15] Granić, A., Nakić, J., Marangunić, N.: Scenario-based group usability testing as a mixed methods approach to the evaluation of three-dimensional virtual learning environments. *Journal of Educational Computing Research*, 58(3), 1–24 (2020).

[PS16] Delamarre, A., Lisetti, C., Buche, C.: A cross-platform classroom training simulator: Interaction design and evaluation. In: *International Conference on Cyberworlds (CW’20)*, 86–93, Caen, France (2020).

[PS17] Birt, J., Vasilevski, N.: Comparison of single and multiuser immersive mobile virtual reality usability in construction education. *Educational Technology & Society*, 24(2), 93–106 (2021).

[PS18] Besoain, F., Jego, L., Gallardo, I.: Developing a virtual museum: Experience from the design and creation process. *Information*, 12(6), 244 (2021).

[PS19] Schmidt, M., Schmidt, C., Glaser, N., Beck, D., Lim, M., Palmer, H.: Evaluation of a spherical video-based virtual reality intervention designed to teach adaptive skills for adults with autism: a preliminary report. *Interactive Learning Environments*, 29(3), 345–364 (2021).

[PS20] Jangid, V., Kongsilp, S.: FishTank sandbox: A software framework for collaborative usability testing of fish tank virtual reality interaction techniques. In: *27th ACM Symposium on Virtual Reality Software and Technology (VRST’21)*, pp. 1–3, article 70, Osaka, Japan (2021).

[PS21] Strada, F., Bottino, A., Lamberti, F., Mormando, G., Ingrassia, P. L.: Holo-BLSD - A holographic tool for self-training and self-evaluation of emergency response skills. *IEEE Transactions on Emerging Topics in Computing*, 9(3), 1581–1595 (2021).

[PS22] Rzeszewski, M., Orylski, M.: Usability of webxr visualizations in urban planning. *ISPRS International Journal of Geo-Information*, 10, 721 (2021).

[PS23] Thalmann, M., Ringli, L., Adcock, M., Swinnen, N., de Jong, J., Dumoulin, C., Guimarães, V., de Bruin, E. D.: Usability study of a multicomponent exergame training for older adults with mobility limitations. *International Journal of Environmental Research and Public Health*, 18, 13422 (2021).

[PS24] Hassandra, M., Galanis, E., Hatzigeorgiadis, A., Goudas, M.; Mouzakidis, C., Karathanasi, E. M., Petridou, N., Tsolaki, M.; Zikas, P.; Evangelou, G., Papagiannakis, G., Bellis, G., Kokkotis, C., Panagiotopoulos, S. R., Giakas, G., Theodorakis, Y.: A virtual reality app for physical and cognitive training of older people with mild cognitive impairment: Mixed methods feasibility study. *JMIR Serious Games*, 9(1), 1–21, article e24170 (2021).

[PS25] Fajar, M., Udjaja, Y., David, Chowanda, A., Juarto, B., Yulianto: A comparative investigation of usability issues toward virtual reality implementation in a state-owned shipping service enterprise. *ICIC Express Letters, Part B: Applications*, 13(5), 545–552 (2022).

[PS26] Kamińska, D., Zwoliński, G., Laska-Leśniewicz, A.: Usability testing of virtual reality applications—The pilot study. *Sensors*, 22, 1342 (2022).

[PS27] Ogunseiju, O. R.; Gonsalves, N.; Akanmu, A. A.; Bairaktarova, D.; Bowman, D. A.; Jazizadeh, F.: Mixed reality environment for learning sensing technology applications in Construction: A usability study. *Advanced Engineering Informatics*, 53, article 101637 (2022).

[PS28] Alpala, L. O.; Quiroga-Parra, D. J.; Torres, J. C.; Peluffo-Ordonez, D. H.: Smart factory using virtual reality and online multi-user: Towards a metaverse for experimental frameworks. *Applied Sciences*, 12, 6258 (2022).

[PS29] Cheng, K. H.: Teachers' perceptions of exploiting immersive virtual field trips for learning in primary education. *Journal of Research on Technology in Education*, 54(3), 438–455 (2022).

[PS30] Mitre-Ortiz, A., Munoz-Arteaga, J., Cardona-Reyes, H.: Developing a model to evaluate and improve user experience with hand motions in virtual reality environments. *Universal Access in the Information Society*, 1–15 (2022).

References

1. Tham, J., Duin, A.H., Gee, L., Ernst, N., Abdelqader, B., McGrath, M.: Understanding virtual reality: presence, embodiment, and professional practice. *IEEE Trans. Prof. Commun.* **61**(2), 178–195 (2018)
2. Kardong-Edgren, S.S., Farra, S.L., Alinier, G., Young, H.M.: A call to unify definitions of virtual reality. *Clin. Simul. Nurs.* **31**, 28–34 (2019)
3. Hickman, L., Akdere, M.: Exploring virtual reality for developing soft-skills in stem education. In: 2017 7th World Engineering Education Forum (WEEF'17), pp. 461–465, Kuala Lumpur, Malaysia (2017)
4. Heiyanthuduwa, T.A., Amarapala, K.W.N.U., Gunathilaka, K.D.V.B., Ravindu, K.S., Wickramarathne, J., Kasthurirathna, D.: VirtualPT: virtual reality based home care physiotherapy rehabilitation for elderly. In: 2020 2nd International Conference on Advancements in Computing (ICAC'20), pp. 311–316, Malabe, Sri Lanka (2020)
5. Li, C., Liang, W., Quigley, C., Zhao, Y., Yu, L.F.: Earthquake safety training through virtual drills. *IEEE Trans. Visual Comput. Graphics* **23**(4), 1275–1284 (2017)
6. Hertzum, M.: Usability testing: A practitioner's guide to evaluating the user experience. *Synthesis Lectures on Human-Centered Informatics*, Morgan & Claypool Publishers (2020)
7. Castro, J.W., Garnica, I., Rojas, L.: Automated tools for usability evaluation: a systematic mapping study. In: Meiselwitz, G. (eds.) *Social Computing and Social Media: Design, User Experience and Impact. HCII 2022. Lecture Notes in Computer Science*, 13315, pp. 28–46. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-05061-9_3
8. Losana, P., Castro, J.W., Ferre, X., Villalba-Mora, E., Acuña, S.T.: A systematic mapping study on integration proposals of the personas technique in agile methodologies. *Sensors* **21**(18), article 6298, Special Issue Recent Advances in Human-Computer Interaction (2021)
9. Ren, R., Zapata, M., Castro, J.W., Dieste, O., Acuña, S.T.: Experimentation for chatbot usability evaluation: a secondary study. *IEEE Access* **10**, 12430–12464 (2022)
10. Barricelli, B.R., De Bonis, A., Di Gaetano, S.: Semiotic framework for virtual reality usability and UX evaluation: a pilot study. In: 2nd Workshop on Games-Human Interaction (GHItaly'18), pp. 1–6, Castiglione della Pescaia, Grosseto, Italy (2018)
11. Moore, A.G., Hu, X., Eubanks, J.C., Aiyaz, A.A., McMahan, R.P.: A formative evaluation methodology for VR training simulations. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW'20), pp. 125–132, Atlanta, GA, USA (2020)

12. Karre, S.A., Mathur, N., Reddy, Y.R.: Understanding usability evaluation setup for VR products in industry: a review study. *ACM SIGAPP Appl. Comput. Rev.* **19**(4), 17–27 (2020)
13. Sheikh, S., Heyat, M.B.B., AlShorman, O., Masadeh, M., Alkhatni, F.: A review of usability evaluation techniques for augmented reality systems in education. In: *2021 Innovation and New Trends in Engineering, Science and Technology Education Conference (IETSEC'21)*, pp. 1–6 (2021)
14. Zanatta, F., Giardini, A., Pierobon, A., D'Addario, M., Steca, P.: A systematic review on the usability of robotic and virtual reality devices in neuromotor rehabilitation: patients' and healthcare professionals' perspective. *BMC Health Serv. Res.* **22**(1), 1–16 (2022)
15. Kitchenham, B.A., Budgen, D., Brereton, O.P.: Using mapping studies as the basis for further research—a participant-observer case study. *Inf. Softw. Technol.* **53**(6), 638–651 (2011)
16. Zhang, H., Babar, M.A., Tell, P.: Identifying relevant studies in software engineering. *Inf. Softw. Technol.* **53**(6), 625–637 (2011)
17. Özding, F., Tüzün, H., Ergün, E., Bayrak, F., Kula, A.: Usability testing of a three-dimensional library orientation game. In: *Games User Research: A Case Study Approach*, Publisher: CRC Press, pp. 77–95 (2016)
18. Pedroli, E., et al.: Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: positive bike. *Sensors (Basel)* **18**(7), 2343 (2018)
19. Oliveira, E., Simões, F.P., Correia, W.F.: Heuristics evaluation and improvements for low-cost virtual reality. In: *2017 19th Symposium on Virtual and Augmented Reality (SVR'17)*, pp. 178–187, Curitiba, Brazil (2017)
20. Naranjo, J.E., Urrutia, F.U., Garcia, M.V., Gallardo-Cárdenas, F., Franklin, T.O., Lozada-Martínez, E.: User experience evaluation of an interactive virtual reality-based system for upper limb rehabilitation. In: *2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG'19)*, pp. 328–333, Quito, Ecuador (2019)
21. Besoain, F., Jego, L., Arenas-Salinas, M.: Implementation of a gamified puzzle based on pro-origami protein structure cartoons: an experience in virtual reality. In: *2018 IEEE Biennial Congress of Argentina (ARGENCON'18)*, pp. 1–7, San Miguel de Tucuman, Argentina (2018)
22. Castro, J.W., Acuña, S.T.: Comparativa de selección de estudios primarios en una revisión sistemática. In *XVI Jornadas de Ingeniería del Software y Bases de Datos (JISBD'11)*, pp. 319–332, A Coruña, España (2011)
23. Magües, D., Castro, J.W., Acuña, S.T.: Usability in agile development: a systematic mapping study. In *XLII Conferencia Latinoamericana de Informática (CLEI'16)*, pp. 677–684, Valpariso, Chile (2016)
24. Ren, R., Castro, J.W., Acuña, S.T., De Lara, J.: Evaluation techniques for chatbot usability: a systematic mapping study. *Int. J. Softw. Eng. Knowl. Eng.* **29** (11n12), 1673–1702 (2019)
25. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **77**(12), 1321–1329 (1994)
26. Çöltekin, A., et al.: Extended reality in spatial sciences: a review of research challenges and future directions. *ISPRS Int. J. Geo Inf.* **9**(7), 439 (2020)
27. Rzeszewski, M., Orylski, M.: Usability of WebXR visualizations in urban planning. *ISPRS Int. J. Geo Inf.* **10**(11), 721 (2021)

Anexo B - Estudios Primarios

En el Anexo, se presenta la Tabla B.1, la cual muestra los 30 estudios primarios obtenidos a partir de la cadena seleccionada y mostrada en la tabla de la Sección 3.1.4. Para cada estudio se proporciona un ID, los nombres de los autores y la referencia correspondiente.

Tabla B.1: Estudios primarios extraídos de la cadena seleccionada

ID	Título	Referencia
1	Am I in the theater? usability study of live performance based virtual reality	[He <i>et al.</i> , 2018]
2	Characteristics, usability, and users experience of a system combining cognitive and physical therapy in a virtual environment: Positive bike	[Pedroli <i>et al.</i> , 2018]
3	Heuristics and Fuzzy Multi-Criteria Decision Making for Evaluating Museum Virtual Tours.	[Kabassi <i>et al.</i> , 2018]
4	Semiotic Framework for Virtual Reality Usability and UX Evaluation: a Pilot Study.	[Barricelli <i>et al.</i> , 2018]
5	Usability evaluation for drone mission planning in virtual reality	[Liu <i>et al.</i> , 2018]
6	Usability evaluation of the interactive 3D virtual reality cultural heritage museum display: Fountain of the lions software application	[Tromp <i>et al.</i> , 2018]

Tabla B.1: Estudios primarios extraídos de la cadena seleccionada

ID	Título	Referencia
7	Usability testing of an interactive virtual reality distraction intervention to reduce procedural pain in children and adolescents with cancer	[Birnie <i>et al.</i> , 2018]
8	Virtual gaming simulation of a mental health assessment: A usability study	[Verkuyl <i>et al.</i> , 2018]
9	An augmented reality approach for ADL support in Alzheimer’s disease: a crossover trial	[Rohrbach <i>et al.</i> , 2019]
10	An interactive virtual training system for assembly and disassembly based on precedence constraints	[Li <i>et al.</i> , 2019]
11	Analysis and comparison of robotics 3d simulators	[De Melo <i>et al.</i> , 2019]
12	Automated usability evaluation of virtual reality applications	[Harms, 2019]
13	Using cognitive walkthrough and hybrid prototyping to gather user requirements in early design virtual reality prototypes	[Doumanis y Economou, 2019]
14	Evaluating and redesigning virtual reality’underwater tourism’application based on heuristic method	[Dantes <i>et al.</i> , 2020]
15	Scenario-based group usability testing as a mixed methods approach to the evaluation of three-dimensional virtual learning environments	[Granić <i>et al.</i> , 2020]
16	A Cross-Platform Classroom Training Simulator: Interaction Design and Evaluation A Cross-Platform Classroom Training Simulator: Interaction Design and Evaluation	[Delamarre <i>et al.</i> , 2020]

Tabla B.1: Estudios primarios extraídos de la cadena seleccionada

ID	Título	Referencia
17	Comparison of single and multiuser immersive mobile virtual reality usability in construction education	[Birt y Vasilevski, 2021]
18	Developing a virtual museum: Experience from the design and creation process	[Besoain <i>et al.</i> , 2021]
19	Evaluation of a spherical video-based virtual reality intervention designed to teach adaptive skills for adults with autism: A preliminary report	[Schmidt <i>et al.</i> , 2021]
20	Fishtank Sandbox: A Software Framework for Collaborative Usability Testing of Fish Tank Virtual Reality Interaction Techniques	[Jangid y Kongsilp, 2021]
21	Holo-BLSD—A Holographic Tool for Self-training and Self-Evaluation of Emergency Response Skills	[Strada <i>et al.</i> , 2019]
22	Usability of WebXR visualizations in urban planning	[Rzeszewski y Orylski, 2021]
23	Usability study of a multicomponent exergame training for older adults with mobility limitations	[Thalmann <i>et al.</i> , 2021]
24	A virtual reality app for physical and cognitive training of older people with mild cognitive impairment: mixed methods feasibility study	[Hassandra <i>et al.</i> , 2021]
25	A comparative investigation of usability issues toward virtual reality implementation in a state-owned shipping service enterprise	[Fajar <i>et al.</i> , 2022]
26	Usability Testing of Virtual Reality Applications—The Pilot Study	[Kamińska <i>et al.</i> , 2022]

Tabla B.1: Estudios primarios extraídos de la cadena seleccionada

ID	Título	Referencia
27	Mixed reality environment for learning sensing technology applications in Construction: A usability study	[Ogunseiju <i>et al.</i> , 2022]
28	Smart factory using virtual reality and online multi-user: Towards a metaverse for experimental frameworks	[Alpala <i>et al.</i> , 2022]
29	Teachers' perceptions of exploiting immersive virtual field trips for learning in primary education	[Cheng, 2022]
30	Developing a model to evaluate and improve user experience with hand motions in virtual reality environments	[Mitre-Ortiz <i>et al.</i> , 2022]