

Evaluación del impacto de QVRLAB: diseño e implementación de pruebas de usuario en un laboratorio virtual de química general

QVRLAB impact evaluation: design and implementation of user tests in a virtual general chemistry laboratory

Juan Gabriel Cabrera Ortiz¹

Cesar David Monroy Rodríguez²

Arbin Iscod Polit Bustos³

Universidad Nacional Abierta y a Distancia, Colombia

Resumen

El presente trabajo describe el diseño y la validación preliminar de un instrumento mixto para evaluar la experiencia de usuario, la usabilidad y la percepción pedagógica del prototipo inmersivo QVRLAB, orientado a prácticas de Química General en la UNAD. El proceso incluyó revisión por expertos (n = 10), entrevistas cognitivas y un pilotaje con estudiantes (n = 30) en sesiones controladas. El cuestionario final consta de 20 ítems agrupados en tres dimensiones: funcionalidad, usabilidad y percepción pedagógica, y utiliza una escala tipo Likert de 5 puntos. La consistencia

¹ Ingeniero de sonido, especialista en Pedagogía para el Desarrollo del Aprendizaje Autónomo, magíster en Ingeniería y doctorando en Gerencia de Proyectos. Líder Nacional del programa Ingeniería Multimedia de la Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI. <https://orcid.org/0000-0002-5213-7813/> juang.cabrera@unad.edu.co

² Ingeniero en multimedia, magister en E-learning y Redes Sociales. Docente del programa Ingeniería en Multimedia de la Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI. <https://orcid.org/0000-0003-2087-2976/> cesard.monroy@unad.edu.co

³ Estudiante del programa de Ingeniería en Multimedia de la Escuela de Ciencias Básicas, Tecnología e Ingeniería – ECBTI. <https://orcid.org/0009-0000-9817-5922/> apolitb@unadvirtual.edu.co

interna observada fue satisfactoria: funcionalidad $\alpha = .82$; usabilidad $\alpha = .85$; percepción pedagógica $\alpha = .88$. El acuerdo entre observadores en la guía de aula, medido con coeficiente de correlación intraclase (CCI), superó 0.75 en los criterios observados. Se combinó análisis cuantitativo (consistencia interna y descriptivos) con análisis cualitativo (observación estructurada y preguntas abiertas) para identificar ajustes en el instrumento y en el prototipo. Con base en los hallazgos, se propone la versión revisada del cuestionario como herramienta para evaluaciones preliminares de UX/usabilidad en entornos inmersivos educativos. Se recomiendan estudios posteriores con muestras ampliadas para confirmar la validez factorial y determinar efectos sobre el aprendizaje mediante diseños pre-post.

Palabras clave: realidad virtual, experiencia de usuario, usabilidad, validación de instrumento, laboratorio virtual, enseñanza de la química.

Abstract

This study describes the design and preliminary validation of a mixed-method instrument to evaluate user experience, usability, and pedagogical perception of the immersive prototype QVRLAB, designed for General Chemistry practices at UNAD. The process included expert review ($n = 10$), cognitive interviews, and a pilot test with students ($n = 30$) in controlled sessions. The final questionnaire consists of 20 items grouped into three dimensions: functionality, usability, and pedagogical perception, using a 5-point Likert scale. Internal consistency was satisfactory: functionality $\alpha = .82$; usability $\alpha = .85$; pedagogical perception $\alpha = .88$. Inter-rater agreement in the classroom observation guide, measured with the Intraclass Correlation Coefficient (ICC), exceeded 0.75 across criteria. Quantitative analyses (internal consistency and descriptives) were combined with qualitative analyses (structured observation and open-ended questions) to identify refinements in both the instrument and the prototype. Based on the findings, the revised questionnaire is proposed as a tool for preliminary UX/usability evaluations in immersive educational environments. Further studies with larger samples are recommended to confirm factorial validity and to assess learning effects through pre-post designs.

Keywords: Augmented reality, user experience, usability, instrument validation, virtual laboratory, chemistry teaching.

1. Introducción

La evolución de las tecnologías inmersivas ha permitido explorar nuevas formas de enseñanza en ciencias básicas. En el caso de la química general, los laboratorios presenciales enfrentan limitaciones recurrentes: costos de operación, disponibilidad de equipos, riesgos de seguridad y barreras de acceso para muchos estudiantes (Merchant *et al.*, 2014; Radianti *et al.*, 2020). QVRLAB se desarrolló con la intención de ofrecer simulaciones controladas que permitan practicar procedimientos de laboratorio —normas de bioseguridad, manejo de materiales volumétricos y no volumétricos, y preparación de soluciones— en un entorno seguro y repetible (Makransky *et al.*, 2019; Mikropoulos & Natsis, 2011).

No obstante, disponer de una interfaz inmersiva técnicamente funcional no garantiza su eficacia pedagógica. Para que los resultados de la implementación sean útiles y transferibles es imprescindible contar con instrumentos de evaluación válidos y fiables que midan la experiencia de usuario, la usabilidad y la percepción pedagógica asociada al prototipo (Cuadros-Vargas *et al.*, 2021). Este artículo se centra, por tanto, en el diseño y la validación preliminar de un instrumento mixto pensado específicamente para QVRLAB, con el objetivo de ofrecer una herramienta reproducible para investigadores y diseñadores instruccionales en ingeniería educativa (Lawshe, 1975; Aiken, 1985).

Este estudio tiene por propósito evaluar, mediante un enfoque mixto y siguiendo principios de validación instrumental, la fiabilidad y pertinencia de un instrumento diseñado para medir la funcionalidad, la usabilidad y la percepción pedagógica del prototipo inmersivo QVRLAB en estudiantes de Química General de la UNAD. Para cumplir este propósito se diseñó un instrumento mixto (cuestionario Likert, guía de observación

y protocolo de entrevista), se sometió a validación cualitativa por expertos y se pilotó con usuarios alfa para análisis psicométrico (Alfa de Cronbach y coeficiente de correlación intraclase —CCI—), complementando con análisis cualitativo para ajustar ítems y formular recomendaciones de mejora (Tavakol & Dennick, 2011; Koo & Li, 2016; Cuadros-Vargas *et al.*, 2021).

2. Reorientado al diseño del instrumento

La evaluación de experiencias de aprendizaje en entornos remotos e inmersivos exige, según Cuadros-Vargas *et al.* (2021), el desarrollo de instrumentos específicos que integren tanto medidas de usabilidad técnica como indicadores pedagógicos relevantes para la enseñanza. En la literatura sobre laboratorios remotos, Merchant *et al.* (2014) y Radianti *et al.* (2020) han documentado que las herramientas de evaluación convencionales (por ejemplo, escalas generales de usabilidad como SUS o UMUX y cuestionarios de presencia) aportan insumos útiles, pero rara vez cubren de manera suficiente las necesidades concretas que plantean los laboratorios virtuales o aumentados; por ello, los autores proponen el diseño de cuestionarios ad hoc que combinen escalas de usabilidad, dimensiones experienciales y componentes de inmersión, seguidos de procesos de reducción y refinamiento mediante análisis psicométrico riguroso (Cuadros-Vargas *et al.*, 2021).

Desde la perspectiva conceptual que sustenta la construcción de ítems, Martín-Gutiérrez *et al.* (2015) y García y Rodríguez (2020) recomiendan distinguir tres dominios teóricos fundamentales que guían la arquitectura del cuestionario: en primer lugar, la usabilidad y la facilidad de uso, entendida como la capacidad del sistema para permitir al usuario ejecutar tareas con eficiencia y sin errores; en segundo lugar, la utilidad o experiencia pedagógica, entendida como la percepción del usuario respecto a la contribución de la herramienta al logro de objetivos de aprendizaje y la comprensión conceptual; y, en tercer lugar, la dimensión de inmersión y motivación, que alude a la sensación de presencia y al grado en que la experiencia sostiene la atención y promueve la participación activa (Mikropoulos & Natsis, 2011; Radianti *et*

al., 2020). La literatura sobre evaluación de experiencias en laboratorios remotos señala que, cuando estas tres dimensiones se articulan de forma coherente en el diseño del instrumento, el resultado es una medición integrada y accionable para diseñadores instruccionales y docentes (Cuadros-Vargas *et al.*, 2021).

Para QVRLAB, y en coherencia con la ficha técnica y los anexos del proyecto, el instrumento se construye además sobre criterios disciplinares, es decir, la alineación explícita con el syllabus de Química General y sobre criterios didácticos derivados de los guiones pedagógicos que definen las tareas críticas a observar en la simulación. Este doble anclaje disciplinar y didáctico, asegura que los ítems no se limiten a capturar sensaciones generales, sino que midan aspectos operativos y pedagógicos concretos y relevantes para la práctica de laboratorio (por ejemplo, comprensión de protocolos, selección adecuada de material y precisión en procedimientos), tal como recomiendan los marcos de validación instrumentales propuestos por Lawshe (1975) y Aiken (1985) para garantizar pertinencia y claridad en juicios de contenido.

En la práctica metodológica seguida en este estudio, la estrategia fue construir primero un banco amplio de ítems que cubriera las tres dimensiones teóricas propuestas, someter ese banco a juicio de expertos para depuración de contenido y comprensibilidad (Lawshe, 1975), y proceder luego a un pilotaje cuyo objetivo es evaluar propiedades psicométricas preliminares, consistencia interna mediante Alfa de Cronbach y acuerdo Inter observador mediante coeficientes de correlación intraclase y, con base en esos resultados, seleccionar y refinar ítems por su desempeño estadístico y su aportación teórica (Tavakol & Dennick, 2011; Koo & Li, 2016). Esta combinación de evidencia cualitativa y cuantitativa permite obtener un instrumento funcional y reproducible para su aplicación en contextos de ingeniería educativa y desarrollo de laboratorios virtuales y realidades aumentadas (Cuadros-Vargas *et al.*, 2021; Radianti *et al.*, 2020).

3. Metodología

3.1 Enfoque y diseño del estudio

Este estudio adopta un enfoque mixto secuencial orientado a la construcción y validación de un instrumento de evaluación para QVRLAB (versión candidata QVRLAB v1.0). Conforme a lo planteado por Cuadros-Vargas *et al.* (2021) en el campo de los laboratorios remotos, el diseño combina fases cualitativas, generación y depuración de ítems, entrevistas cognitivas y juicio de expertos para validación de contenido, con fases cuantitativas, pilotaje con usuarios y análisis psicométrico preliminar (consistencia interna y acuerdo inter-evaluador). La secuencia metodológica que guía el estudio asegura trazabilidad y reproducibilidad:

Generación del banco de ítems → depuración lingüística mediante entrevistas cognitivas → validación de contenido por expertos → pilotaje alfa y análisis psicométrico preliminar → recomendaciones y plan de uso futuro (Radianti *et al.*, 2020; Cuadros-Vargas *et al.*, 2021).

3.2 Participantes

El panel de validación de contenido estará conformado por diez expertos ($n = 10$) integrados por docentes de Química General, especialistas en diseño instruccional y profesionales en experiencia de usuario (UX), tal como recomiendan las guías para juicios de contenido (Lawshe, 1975; Aiken, 1985).

La fase de revisión profesional (pilotaje alfa) contará con diez participantes disciplinares ($n = 10$; docentes/usuarios con experiencia en prácticas de laboratorio) cuya función es evaluar pertinencia y funcionalidad desde la perspectiva pedagógica. Finalmente, la prueba piloto con usuarios implicará una muestra de treinta estudiantes ($n = 30$) de Química General (usuarios alfa). Los criterios de inclusión requieren ser mayor de 18 años y estar matriculado en asignaturas básicas de química; entre los criterios de exclusión preventiva se incluye diagnóstico de epilepsia fotosensible u otras condiciones que contraindiquen la exposición a estímulos visuales intensos, conforme a las consideraciones éticas y de seguridad consignadas en la ficha del proyecto y en el Formato Anexo

3.3 Instrumentos y materiales

Para esta investigación se diseñó y empleó un conjunto articulado de instrumentos, descritos a continuación y anclados a los guiones pedagógicos del prototipo.

En primer lugar, el *instrumento diseñado QVRLAB v1.0*: un cuestionario tipo Likert de 5 puntos (1 = Totalmente en desacuerdo; 5 = Totalmente de acuerdo) compuesto por 20 ítems distribuidos en tres dimensiones teóricas —Funcionalidad (7 ítems), Usabilidad (7 ítems) y Percepción pedagógica (6 ítems)— e incluyendo dos ítems invertidos para control de sesgo. La instrucción administrativa indica que cada participante evaluará un único escenario del prototipo (Interfaz introductoria; Laboratorio + tablet; Reto 3) y responderá en función de esa experiencia para evitar mezclas de vivencias.

En segundo lugar, se empleó una *guía de observación estructurada* con indicadores operativos (p. ej., selección de material, cumplimiento de la secuencia procedimental, precisión en mediciones, uso y aprovechamiento de la retroalimentación, y signos conductuales de frustración o confusión), aplicada por dos observadores independientes para estimar acuerdo inter-evaluador.

En tercer lugar, se diseñó una *entrevista semiestructurada breve* (5–10 preguntas) para recuperar impresiones cualitativas sobre la claridad de los ítems, barreras de comprensión e ideas de mejora.

Finalmente, se utilizó una *plantilla de registro* para la aplicación (ParticipantID, Grupo: estudiante/docente/experto, Escenario evaluado, timestamps, observadores, respuestas por ítem y comentarios); la plantilla completa está disponible en los anexos y en el archivo QVRLAB_ItemMap_and_ResponseTemplate (Cuadros-Vargas *et al.*, 2021).

3.4 Procedimiento (paso a paso)

La intervención metodológica siguió los pasos que se describen en el guion operativo del estudio y que permiten reproducir la recolección y la validación instrumental.

En primera instancia se realizó la *generación del banco de ítems* a partir del marco teórico, los objetivos instruccionales y el análisis de los guiones pedagógicos del prototipo; cada ítem fue anclado a una conducta observable o a una evidencia operativa del guion.

Posteriormente se llevaron a cabo *entrevistas cognitivas* ($n = 4-6$) con estudiantes para detectar ambigüedades semánticas y verificar la comprensibilidad de los enunciados; los ítems problemáticos fueron reformulados.

A continuación, la *validación por expertos* ($n = 10$) evaluó pertinencia y claridad mediante una ficha de juicio que permitió calcular indicadores de validez de contenido (I-CVI por ítem; S-CVI para la escala) y recoger observaciones cualitativas (Lawshe, 1975; Aiken, 1985).

Tras incorporar las sugerencias del panel, se realizó un *pilotaje alfa con docentes* ($n = 10$) para una última revisión profesional de redacción y pertinencia; finalmente, el *pilotaje con estudiantes* ($n = 30$) consistió en sesiones controladas en RA (tablet/móvil) en las cuales cada participante interactuó con un solo escenario, fue observado por dos evaluadores, completó el instrumento diseñado y participó en la entrevista semiestructurada breve. Todas las sesiones siguieron un protocolo estandarizado que incluye instrucciones claras, tiempo estimado de aplicación y criterios de suspensión en caso de malestar. Las versiones del instrumento, actas de expertos y registros de entrevistas cognitivas y observaciones se consignaron en anexos para asegurar trazabilidad metodológica (Radianti *et al.*, 2020; Cuadros *et al.*, 2021).

3.5 Estrategia de aplicación multiactor y por escenario

Siguiendo recomendaciones para evaluar UX y usabilidad en contextos educativos, el instrumento se concibió como multiactor — aplicable a estudiantes, docentes y expertos, por escenario cada participante evalúa únicamente uno de los tres entornos del prototipo; la misma batería de 20 ítems se administra a los tres perfiles con el propósito de posibilitar comparaciones entre actores y detectar diferencias en la sensibilidad de ítems según el perfil profesional o formativo. La administración individual evita sesgos por discusión en grupo y facilita análisis matriciales (grupo $\times$ escenario $\times$ ítem) que permitirán, en fases posteriores, pruebas de funcionamiento diferencial (DIF) entre subpoblaciones (Cuadros-Vargas *et al.*, 2021).

3.6 Análisis de datos

la dimensión cualitativa, las entrevistas semiestructuradas y las notas de observación serán sometidas a codificación temática para identificar problemas de comprensión, propuestas de mejora y evidencia cualitativa de validez de contenido; los juicios de expertos se sintetizarán y se calcularán índices de acuerdo de contenido (I-CVI por ítem; S-CVI para la escala) como criterios de retención o reformulación de ítems (Lawshe, 1975). En la dimensión cuantitativa preliminar, se calculará la consistencia interna por dimensión mediante el Alfa de Cronbach (Tavakol & Dennick, 2011) y se estimará el coeficiente de correlación intraclase (ICC/CCI) para las observaciones de los dos evaluadores (Koo & Li, 2016), seleccionando el modelo apropiado (por ejemplo, ICC(2,1) para medidas interevaluador con efectos aleatorios). Adicionalmente se examinarán distribución de respuestas por ítem, discriminación ítem-total y contribución ítem-escala; ítems con baja contribución serán revisados o descartados según criterios estadísticos y de contenido. Con muestras ampliadas se planifican análisis factorial exploratorio (EFA) y análisis factorial confirmatorio (CFA), así como test-retest para estimar estabilidad temporal y análisis DIF entre grupos y escenarios.

3.7 Criterios de decisión y corte

Los criterios de decisión metodológica combinan evidencia de contenido y evidencia empírica: i) indicadores de validez de contenido (I-CVI / S-CVI y, cuando proceda, Aiken V) guiarán la retención o reformulación de ítems; ii) valores de Alfa de Cronbach por dimensión inferiores a 0.70 motivarán revisión de ítems ($\alpha \geq 0.70$ se considerará aceptable en esta etapa preliminar) (Tavakol & Dennick, 2011); iii) para ICC se empleará el marco interpretativo recomendado por Koo y Li (2016): < 0.50 (pobre), $0.50-0.75$ (moderado), $0.75-0.90$ (bueno) y > 0.90 (excelente). Estas reglas combinadas permiten decisiones fundamentadas tanto estadística como conceptualmente.

3.8 Consideraciones éticas y logísticas

Dado que las pruebas de validación descritas en este estudio se realizan empleando gafas de Realidad Virtual (HMD), el diseño del

instrumento integra de manera explícita las medidas necesarias para documentar la *seguridad y tolerancia* de la exposición inmersiva como parte misma de la evaluación de usabilidad. Radianti *et al.* (2020) y Parong & Mayer (2018) subrayan que, en entornos inmersivos, es tanto necesario como práctico incorporar en la fase de diseño elementos que permitan monitorizar síntomas físicos y cognitivos vinculados a la exposición. En coherencia con ello, el Instrumento QVRLAB v1.0 incluye módulos específicos y metadatos que recogen: (a) el informe autoaplicado sobre síntomas (mareo, náusea, visión borrosa, fatiga visual) aplicado antes, durante y después de la sesión; (b) la duración efectiva de la exposición y la contabilización de pausas; (c) la evidencia de activación del mecanismo de parada o interrupción de la experiencia; y (d) registros técnicos (modelo de HMD, tiempo de latencia y ajustes principales) que permitan relacionar experiencias de uso con características del hardware. Estos elementos se complementan con una *guía de aplicación* que especifica cómo recoger y codificar esa información (Cuadros-Vargas *et al.*, 2021).

Asimismo, el instrumento incorpora una plantilla de observación para los evaluadores, destinada a documentar indicadores conductuales (desorientación, signos de incomodidad, necesidad de asistencia) que sirven de triangulación frente a los auto-informes. En lo ético, la aplicación del instrumento se realiza bajo el amparo del consentimiento informado consignado en el Anexo 9, donde se explican los objetivos del estudio, la naturaleza voluntaria de la participación y las medidas de protección de datos personales. Dado que el interés central de este artículo es el *diseño y la validación del instrumento*, las decisiones operativas detalladas sobre reclutamiento, protocolos técnicos y logística de aplicación quedarán formalizadas por el equipo desarrollador y sometidas a los procedimientos institucionales correspondientes; el instrumento, sin embargo, incluye las herramientas de medición necesarias para que esas decisiones puedan evaluarse empíricamente en fases ulteriores (Radianti *et al.*, 2020; Parong & Mayer, 2018).

3.9. Limitaciones metodológicas previstas

El empleo de HMD en la validación del instrumento introduce limitaciones metodológicas que es imprescindible reconocer desde el diseño. En primer

lugar, la variabilidad de tolerancia individual a experiencias inmersivas (incluida la cinetosis o cybersickness) puede condicionar tanto las respuestas de usabilidad como el rendimiento en tareas procedimentales; por ello, aunque el instrumento registra síntomas y tiempos de exposición, la medición por sí sola no elimina la posibilidad de que el malestar influya en los resultados y, por tanto, la interpretación debe considerar esta fuente de sesgo (Radianti *et al.*, 2020). En segundo lugar, la heterogeneidad del hardware (modelo de HMD, resolución, latencia, ergonomía) y de las condiciones ambientales puede modular la experiencia de usuario; aun cuando el instrumento recoge metadatos de dispositivo, la generalización de los hallazgos a otros equipos precisará validaciones complementarias con distintas configuraciones (Merchant *et al.*, 2014).

Metodológicamente, la naturaleza principalmente auto-reportada de algunas dimensiones (experiencia, percepción pedagógica, molestias) puede verse afectada por sesgos de respuesta; la estrategia multiactor y la triangulación con observación estructurada mitigan, pero no eliminan, este problema. Además, los indicadores psicométricos obtenidos con un pilotaje preliminar ($n = 30$) proporcionan estimaciones iniciales de consistencia interna y acuerdo inter-evaluador, pero no garantizan la estabilidad factorial ni la validez convergente definitiva; por tanto, como señalan Tavakol & Dennick (2011) y Koo & Li (2016), se requieren muestras ampliadas y análisis adicionales (EFA/CFA, test-retest) para confirmar la estructura y la fiabilidad del instrumento en diferentes poblaciones y contextos técnicos.

Finalmente, es importante destacar que muchas de las decisiones de mitigación del riesgo (por ejemplo, límites de duración de exposición, pausas obligatorias, criterios de suspensión) están reflejadas en la propia batería y en las guías de aplicación del instrumento. Por consiguiente, las conclusiones de este trabajo deben entenderse como aplicables al instrumento y al protocolo de aplicación específicos descritos aquí y no necesariamente extensibles a otras implementaciones sin una validación previa adaptada al nuevo hardware, guiones pedagógicos o población objetivo.

4. Resultados

Los resultados de esta etapa se concentran exclusivamente en el producto instrumental: la versión candidata del instrumento diseñado QVRLAB v1.0, su estructura y las decisiones metodológicas adoptadas tras la revisión de expertos y el pilotaje alfa. Tras la depuración y la validación de contenido, la versión candidata quedó compuesta por veinte (20) ítems organizados en tres dimensiones conceptuales claramente definidas: Funcionalidad (siete ítems), Usabilidad (siete ítems) y Percepción pedagógica (seis ítems). Los ítems fueron redactados y anclados a conductas y evidencias concretas derivadas de los guiones pedagógicos del prototipo, de modo que la interpretación de las puntuaciones resulte directamente accionable para el equipo de diseño instruccional y permita detectar problemas operativos y formativos en cada escenario (Cuadros-Vargas *et al.*, 2021).

En las rondas de revisión cualitativa con el panel de expertos se tomó una decisión operativa relevante para la aplicabilidad del instrumento: se acordó aplicar la misma batería de ítems a tres perfiles de evaluadores —estudiantes, docentes y expertos— y administrar el instrumento por escenario, de forma que cada participante evalúe únicamente uno de los tres entornos del prototipo (Interfaz introductoria y navegación inicial; Laboratorio y uso de la tablet; Navegación en el reto práctico). Esta decisión, sugerida por los evaluadores, persigue dos objetivos complementarios: obtener evidencia comparativa entre actores sobre la sensibilidad e interpretación de los ítems, y garantizar que las respuestas reflejen la vivencia concreta de una tarea definida por el guion pedagógico (Lawshe, 1975; Cuadros-Vargas *et al.*, 2021).

El pilotaje alfa permitió detectar y corregir ambigüedades lingüísticas, reformular ítems de bajo entendimiento y eliminar redundancias; asimismo, la observación estructurada aportó evidencia de situaciones procedimentales que requerían ítems adicionales de anclaje operativo. En particular, se documentaron casos en los que la precisión en la selección de materiales y la lectura de escalas exigían enunciados más precisos, por lo que varios ítems fueron reescritos para mejorar su relación ítem-evidencia. Todas las actas de revisión, las versiones previas y posteriores de los ítems y los registros del pilotaje están consignados en los anexos para garantizar trazabilidad y permitir re-evaluaciones futuras (Radianti *et al.*, 2020).

En cuanto a la dimensión psicométrica, esta etapa entregó diagnósticos preliminares y procedimientos de análisis planificados. La

estructura multiactor y multi-escenario diseñada facilita la generación de matrices de datos (grupo $\times$ escenario $\times$ ítem) que permitirán, en fases subsiguientes, realizar análisis de consistencia interna por dimensión (Alfa de Cronbach), estimaciones de acuerdo inter-evaluador (CCI/ICC) y, con muestras ampliadas, análisis factoriales (EFA/CFA) y pruebas de funcionamiento diferencial de ítems (DIF). En síntesis, el resultado tangible de esta fase es un instrumento candidato, técnicamente anclado y operativo, listo para su aplicación en estudios ampliados y para la obtención de evidencias psicométricas robustas (Tavakol & Dennick, 2011; Koo & Li, 2016).

5. Conclusiones

El trabajo conducido permitió desarrollar y dejar como versión candidata el instrumento diseñado QVRLAB v1.0, compuesto por veinte ítems organizados en tres dimensiones conceptuales (Funcionalidad, Usabilidad y Percepción pedagógica), acompañado de una guía de observación y de un protocolo de entrevista semiestructurada. La redacción de los ítems fue deliberadamente anclada a los guiones pedagógicos del prototipo con el fin de asegurar la interpretabilidad operativa de las respuestas y la utilidad inmediata para el equipo de diseño instruccional.

La validación cualitativa realizada, a través de entrevistas cognitivas y del juicio de un panel de expertos, permitió depurar la redacción, eliminar ambigüedades y ajustar la cobertura de los ítems respecto a los objetivos pedagógicos y las exigencias de usabilidad. El pilotaje alfa, complementado con observación estructurada, proporcionó evidencia práctica que orientó la reformulación de ítems problemáticos y la confirmación de la estrategia multiactor y por escenario. Toda la evidencia documental generada en estas etapas se encuentra registrada en anexos, garantizando trazabilidad metodológica y posibilidad de auditoría.

En la presente etapa se completó la fase de diseño y validación de contenido; las acciones instrumentales necesarias para culminar la validación psicométrica, cálculo de Alfa de Cronbach por dimensión, estimación del CCI/ICC, análisis factorial exploratorio y confirmatorio y test-retest, han sido planificadas como pasos siguientes y se recomienda ejecutarlas con muestras ampliadas para obtener estimaciones robustas de consistencia, estructura factorial y estabilidad temporal. Hasta tanto

se cuente con esos datos ampliados, las evidencias psicométricas se deben considerar preliminares.

Finalmente, el instrumento diseñado QVRLAB v1.0 se presenta como una versión candidata válida para evaluaciones preliminares y como una herramienta reproducible para investigadores y diseñadores instruccionales interesados en valorar la usabilidad y la percepción pedagógica de laboratorios inmersivos. Su configuración multiactor y por escenario, así como la inclusión de ítems anclados a evidencias operativas, facilitan su aplicabilidad práctica y su adaptación futura a estudios comparativos, análisis factoriales y evaluaciones de funcionamiento diferencial entre subpoblaciones.

Referencias

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142.
- Alqahtani, M., & Daghestani, L. (2020). Immersive virtual reality for learning: A review. *Education and Information Technologies*, 25, 5759–5789. <https://doi.org/10.1007/s10639-020-10324-5>
- Cabero-Almenara, J., & Barroso-Osuna, J. (2019). Los escenarios de aprendizaje basados en realidad aumentada: posibilidades y retos. *Comunicar*, 27(59), 111–120. <https://doi.org/10.3916/C59-2019-10>
- Calderón, S. J., Tumino, M. C., & Bournissen, J. M. (2020). Realidad virtual: impacto en el aprendizaje percibido de estudiantes de Ciencias de la Salud. *Tecnología, Ciencia y Educación*, 16, 65–82. <https://doi.org/10.51302/tce.2020.441>
- Cantón Enríquez, D., Arellano Pimentel, J. J., Hernández López, M. Á., & Nieva García, O. S. (2017). Uso didáctico de la realidad virtual inmersiva con interacción natural de usuario enfocada a

la inspección de aerogeneradores. *Apertura*, 9(2), 8–23.
<https://doi.org/10.18381/Ap.v9n2.1049>

Cuadros-Vargas, E., Serrano, V., García-Zubía, J., & Hernández-Jayo, U. (2021). Design and evaluation of a user experience questionnaire for remote labs. *IEEE Transactions on Education*, 64(2), 115–122. <https://doi.org/10.1109/TE.2020.3032014>

Domínguez, A., & Jaime, A. (2018). Aprendizaje inmersivo: un estudio sobre la motivación en laboratorios virtuales. *RIED. Revista Iberoamericana de Educación a Distancia*, 21(1), 93–110. <https://doi.org/10.5944/ried.21.1.18837>

Fernández-Batanero, J. M., & Bermejo, B. (2020). Realidad virtual en educación superior: estudio de caso sobre usabilidad. *RED. Revista de Educación a Distancia*, 20(64), 1–20. <https://doi.org/10.6018/red.402951>

Fombona, J., Pascual-Sevillano, M. Á., & González-Videgaray, M. (2017). Realidad virtual y educación: posibilidades y retos. *Pixel-Bit. Revista de Medios y Educación*, 50, 197–210. <https://doi.org/10.12795/pixelbit.2017.i50.13>

Fonseca, D., Redondo, E., & Valls, F. (2016). Motivación y mejora del aprendizaje en realidad aumentada: validación de un cuestionario. *Computers in Human Behavior*, 57, 604–612. <https://doi.org/10.1016/j.chb.2015.01.036>

García, F., & Rodríguez, M. (2020). Estrategias de evaluación de la experiencia de usuario en entornos inmersivos. *Revista de Educación Mediática y TIC (EDMETIC)*, 9(2), 72–91. <https://doi.org/10.21071/edmetic.v9i2.12063>

González, C. S. (2019). Usabilidad en entornos educativos: revisión y propuestas de evaluación. *RIED. Revista Iberoamericana de Educación a Distancia*, 22(2), 245–264. <https://doi.org/10.5944/ried.22.2.23239>

- Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23, 1515–1529. <https://doi.org/10.1007/s10639-017-9676-0>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575.
- Lerma García, L., Rivas Porras, D., Adame Gallegos, J. R., Ledezma Millán, F., López De La Torre, H. A., & Ortiz Palomino, C. E. (2020). Realidad Virtual como técnica de enseñanza en Educación Superior: perspectiva del usuario. *Enseñanza & Teaching: Revista Interuniversitaria de Didáctica*, 38(1), 111–123. <https://doi.org/10.14201/et2020381111123>
- Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236. <https://doi.org/10.1016/j.learninstruc.2017.12.007>
- Martín-Gutiérrez, J., Fabiani, P., Benesova, W., Meneses, M. D., & Mora, C. E. (2015). Augmented reality to promote collaborative and autonomous learning in higher education. *Computers in Human Behavior*, 51, 752–761. <https://doi.org/10.1016/j.chb.2014.11.093>
- Mejías Martínez, G., Cuesta Díaz, V., & González Vallés, J. E. (2024). Realidad virtual, aumentada y mixta en el sector salud: perspectivas y experiencia del usuario en educación superior universitaria. *European Public & Social Innovation Review*, 9, 1–14. <https://doi.org/10.31637/epsir-2024-558>
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based

- instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40. <https://doi.org/10.1016/j.compedu.2013.07.033>
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research. *Computers & Education*, 56(3), 769–780. <https://doi.org/10.1016/j.compedu.2010.10.020>
- Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, 10(6), 549–559. <https://doi.org/10.1002/ase.1696>
- Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785–797. <https://doi.org/10.1037/edu0000241>
- Paul, R., & Elder, L. (2008). *The miniature guide to critical thinking concepts and tools* (rev. ed.). Foundation for Critical Thinking.
- Predescu, S. L., Caramihai, S. I., & Moisescu, M. A. (2024). Percepciones de autoeficacia en docentes en formación en España: un estudio de caso de microenseñanzas utilizando realidad virtual inmersiva. *Pixel-Bit. Revista de Medios y Educación*, (71), 7–24. <https://doi.org/10.12795/pixelbit.107712>
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Romero-Esquinas, M. H., Hidalgo-Ariza, M. D., Muñoz-González, J. M., & Ariza-Carrasco, C. (2024). La realidad virtual y el Diseño Universal de Aprendizaje: una manera inclusiva y actual de entender la educación. *Revista de Investigación Educativa*, 42(1), 123–140. <https://doi.org/10.6018/rie.564881>

- Rubio-Tamayo, J. L., Gertrudix Barrio, M., & García García, F. (2018). Experiencia de usuario y diseño de interacciones en procesos creativos y ciencias educativas con tecnologías de realidad virtual y aumentada. *Enseñanza & Teaching: Revista Interuniversitaria de Didáctica*, 36(1), 63–79. <https://doi.org/10.14201/et2183616379>
- Sandoval, J., & Tabash, H. (2021). Realidad virtual como apoyo innovador en la educación a distancia. *Innovaciones Educativas*, 23. <https://doi.org/10.22458/ie.v23iespecial.3622>
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, Article 74. <https://doi.org/10.3389/frobt.2016.00074>
- Tamayo Ravilla, J. R., & Acosta Rosales, R. A. (2024). Diseño + VR – Revisión Teórica de Conceptos. Revisión conceptual de la Experiencia de Usuario en Realidad Aumentada y Realidad Virtual. *Cuadernos del Centro de Estudios de Diseño y Comunicación*, (225). <https://doi.org/10.18682/cdc.vi225.11248>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Verdugo Guamán, G., & Ramón Pacurucu, P. (2024). Realidad virtual en 3D: una herramienta inclusiva para la educación. *Mamakuna*, (22), 76–87. <https://doi.org/10.70141/mamakuna.22.878>
- Villalustre, L., & Del Moral, M. E. (2017). Realidad virtual inmersiva para la enseñanza de ciencias: diseño y experimentación de un laboratorio virtual. *Educación XX1*, 20(2), 93–116. <https://doi.org/10.5944/educxx1.19033>

