

Diseño y validación de contenido de una rúbrica para medir el valor educativo de aplicaciones para dispositivos móviles

Design and content validity of a rubric to measure the educational value of applications for mobile devices

Jorge Expósito López
Universidad de Granada. Granada, España
jorgeel@ugr.es

José Javier Romero Díaz de la Guardia
Universidad de Granada. Granada, España
jjromero@ugr.es

Eva María Olmedo Moreno
Universidad de Granada. Granada, España
emolmedo@ugr.es

Resumen

Las aplicaciones para dispositivos móviles son cada vez más un recurso frecuentemente utilizado en los diferentes escenarios educativos. El objetivo de este trabajo fue el diseño y validación de un instrumento para medir el valor educativo de estas aplicaciones, al que se ha denominado E-V-A (Educational Value Assessment). Se siguió un proceso estructurado en tres fases, acorde a una metodología mixta, contemplando el diseño conceptual y teórico de la rúbrica, diseño funcional de dimensiones, indicadores y descriptores de valoración y posterior validación de contenido teórico-práctica mediante juicio de expertos en torno a la pertinencia, claridad y adecuación, utilizando técnica Delphi. El proceso de validación permitió la formulación de una rúbrica de evaluación de app educativas compuesta por 100 ítems estructurados en 21 bloques y 5 dimensiones: desarrollo personal, desarrollo cognitivo y metacognitivo, desarrollo del aprendizaje y de competencias; desarrollo social y dimensión técnica. Este instrumento puede ser de gran utilidad para profesionales de la docencia que necesiten medir el valor educativo de una app mediante un protocolo estructurado y uniforme.

Palabras clave: aprendizaje móvil; aplicación educativa; evaluación; valor educativo; rúbrica; validación

Abstract

Applications for mobile devices are increasingly a frequently used resource in different educational settings. The objective of this work was the design and validation of an instrument to measure the educational value of these applications, which has been called E-V-A (Educational Value Assessment). A process comprised of three phases was followed, according to a mixed methodology, considering the conceptual and theoretical design of the rubric, the functional design of dimensions, indicators and descriptors, as well as the subsequent validation of theoretical-practical content through expert judgment on relevance, clarity and adequacy, using the Delphi technique. The validation process allowed the formulation of an evaluation rubric for educational apps made up of 100 items structured into 21 blocks and 5 dimensions: personal development, cognitive and metacognitive development, learning and skills development; social development and technical dimension. This instrument can be very useful for teaching professionals who need to measure the educational value of an app through a structured and uniform protocol.

Key words: mobile learning; educational application; app; evaluation; educational value; rubric; validation

1. Introducción

Las aplicaciones para dispositivos móviles (Apps) han alcanzado una gran relevancia en los entornos digitales actuales y se consideran herramientas que pueden favorecer la transformación sociocultural en diferentes planos, desde la salud, las relaciones interpersonales hasta el entretenimiento, las finanzas y también en escenarios educativos. Según Brazuelo-Grund et al. (2017, p. 12) “la vida ya es móvil y las instituciones educativas, en todos los niveles y ámbitos, deberían esforzarse por integrar y normalizar, y no por excluir sin más, estas potenciales herramientas educativas”.

El interés de la propuesta planteada se contextualiza en una realidad palpable en la nueva ciudadanía digital y sus nuevas formas de interacción que emplean masivamente herramientas digitales masivas como las Apps. De acuerdo con las ideas de Vygotsky y sus continuadores, los procesos psicológicos superiores se caracterizan por la utilización de instrumentos de origen cultural adquiridos socialmente, particularmente instrumentos simbólicos como el lenguaje u otros sistemas de representación. Este uso permite la adaptación activa al medio, así como la realización y el control consciente que caracterizan dichos procesos psicológicos superiores. En tal escenario, el uso de las herramientas digitales, su lenguaje y simbología determinará el propio proceso cognitivo, si bien precisará la adquisición de competencias de pensamiento de orden superior y competencias digitales (Almerich et al., 2018).

Desde esta perspectiva, las Apps constituyen un medio de representación y comunicación novedoso, cuyo uso puede introducir modificaciones importantes en determinados aspectos del funcionamiento psicológico de las personas. Un medio que crea, a partir de la integración de tales sistemas, condiciones totalmente nuevas de tratamiento, transmisión, acceso y uso de la información (Morales et al., 2020). Aunque en sí mismo esta realidad no constituye en sentido estricto un nuevo sistema semiótico dado que se emplean sistemas previamente existentes, como el lenguaje oral y escrito, la imagen audiovisual o las representaciones gráficas.

La relevancia del uso creciente y masivo de estas aplicaciones puede constatararse en la existencia de una regulación normativa de la Unión Europea, Directiva (UE) 2016/2102 del 26 de octubre de 2016 (European Parliament, 2016), que establece las necesidades de accesibilidad de los sitios web y las aplicaciones para dispositivos móviles de los organismos del sector público para todos los ciudadanos. La existencia de una reglamentación sobre la obligatoriedad de garantizar la accesibilidad de los espacios virtuales y las Apps, pone de manifiesto que este tipo de recursos se van a utilizar de forma masiva y que se trata elementos consustanciales en la sociedad actual.

De forma habitual, los adolescentes emplean las tecnologías, los dispositivos móviles y Apps, tanto ámbitos formales como informales, para gestionar, aprender y divertirse (Symeonaki et al., 2022). Este uso ha dado lugar a nuevas formas de comunicarse, relacionarse e interactuar, generando una ciudadanía y una sociedad digital. Además, los sistemas educativos formales y las familias están depositando una creciente responsabilidad formativa en el uso de dichas aplicaciones. Wai et al (2018) enfatizan

diferentes beneficios derivados del uso de Apps en contextos educativos, como la portabilidad, la flexibilidad, así como su capacidad para romper las restricciones del tiempo y el espacio físico, lo que lleva a la ubicuidad del aprendizaje, cuestión esta última que ya analizaba Zapata-Ros (2012) poniendo de manifiesto la necesaria evaluación de la solidez pedagógica en contextos de aprendizaje móvil. Igualmente, es destacable el poder motivacional de las Apps como utilidad percibida en el plano educativo (Camilleri & Camilleri, 2019; Li et al., 2018; Morales et al., 2020).

Además del posible efecto motivador de los dispositivos móviles en el aula, también es importante considerar la actitud del profesorado ante el uso de estos recursos. En un estudio basado en percepciones del profesorado en torno al uso que se hace del móvil con fines educativos, Tabuenca et al. (2019) destacaron que los docentes muestran una predisposición al empleo de dispositivos móviles como herramienta de apoyo a los procesos de tutorización del alumnado.

Las ventajas aludidas y la amplia diversidad y disponibilidad de Apps, pueden inducir a los profesionales docentes a utilizarlas sin prestar suficiente atención a la necesaria reflexión previa y la constatación de su idoneidad para el escenario de enseñanza-aprendizaje en el que se deseen integrar. Generalmente, la primera referencia que se toma está basada en la satisfacción que los usuarios o desarrolladores otorgan en forma de estrellas (puntuación de 1 a 5), que se hace pública desde los principales centros de distribución de Apps. Estas evaluaciones pueden resultar engañosas porque están más influenciadas por la usabilidad y la funcionalidad que por la calidad del contenido. Además, no existe garantía a priori de que las aplicaciones clasificadas como educativas sean realmente útiles para el contexto en el que van a utilizarse, ya que no es obligatorio para desarrollares hacer pública tal información (Taylor et al., 2022).

Por otro lado, en la mayoría de los casos, cuando se realiza una evaluación algo más profunda se hace empleando criterios cuasi-tecnológicos o meramente técnicos. Lo que supone un procedimiento poco válido y fiable para conocer su potencial y uso educativo. A pesar de que muchos desarrolladores de software son conscientes de la importancia de consultar con profesionales del sector de la educación, se sigue creando software sin la consideración pertinente de aquellos factores educativos clave que pueden afectar al aprendizaje. De acuerdo con Stoyanov et al. (2015), esto es debido a la escasa disponibilidad de la información sobre la calidad de las Apps más allá de las clasificaciones publicadas en las páginas web de las tiendas y los comentarios de los usuarios. Es decir, las valoraciones públicas de las Apps son subjetivas y pueden originarse en fuentes poco fiables. En este sentido, Lee y Cherner (2015), plantean la importancia de disponer de protocolos que permitan evaluar el potencial educativo de las Apps dada la ausencia de información sobre esta cuestión en los entornos de descarga de aplicaciones.

En los últimos años, se ha intentado realizar evaluaciones de Apps mediante el uso de rúbricas, aunque con limitado éxito. Así por ejemplo, en el ámbito de Apps destinadas a alumnado de preescolar, no se han localizado demasiadas rúbricas o estas han sido inadecuadas e inespecíficas como indican Papadakis et al., (2017). Uno de los grandes problemas en estos procesos de evaluación, según Cherner et al. (2016), es que los

criterios de evaluación utilizados en la mayoría de las rúbricas no están bien vinculados con la investigación previa. En consecuencia, las dimensiones de evaluación de estos instrumentos pueden no estar definidas de forma rigurosa y, por tanto, los indicadores que derivan de ellas pueden ser ineficaces.

La revisión de algunos de los estudios sobre la construcción y uso de escalas evaluativas de Apps educativas revelan sus limitaciones. Por ejemplo, Chau (2014) creó una escala o formulario, conocido como “Developmentally Appropriate App Design Evaluation”, para evaluar que el diseño fuese apropiado al desarrollo de los usuarios en cuatro principios básicos: interacción, aspecto visual, información acústica y diseño instructivo. En este caso se considera el aspecto instructivo al mismo nivel que aspectos funcionales de interacción y físico-perceptivo, cuando deberían plantearse solamente como un elemento vehicular para alcanzar este último.

En un estudio sobre desarrollo de una rúbrica de evaluación de Apps para dispositivos móviles en el contexto de la educación especial, Weng (2015) estableció como categorías de evaluación de aplicaciones: las características de diseño, las posibilidades de individualización, el contenido, la usabilidad, la provisión de feedback y la evaluación. Por otro lado, Walker (2011) consideró que los ejes esenciales que deben guiar los procesos de evaluación de Apps deben ser la conexión con el currículum, la eficacia con la que las Apps integran experiencias de aprendizaje auténticas, las posibilidades de provisión de feedback, la capacidad de establecer niveles de dificultad o centrarse en habilidades específicas, la facilidad de uso y la motivación.

La mayoría de propuestas de evaluación otorgan una mayor importancia a la evaluación de criterios técnicos, como Shoukry et al. (2015) que crearon un marco de evaluación en un intento de investigar la efectividad y propiedad de los juegos educativos, planteando para ello un conjunto de categorías que consideraban aspectos como el diseño de pantalla, la navegación y control de la App, la usabilidad, la presentación del contenido, los problemas de seguridad, accesibilidad y costes. Otros enfoques incorporan dimensiones más cercanas a los procesos cognitivos y de aprendizaje como el rigor, el ejercicio de habilidades del siglo XXI, la conexión con futuros aprendizajes, el valor de los errores, o el aprendizaje cooperativo entre otras (Lee & Cherner, 2015).

Papadakis et al. (2017), realizan una amplia revisión bibliográfica sobre algunas escalas de evaluación de Apps y concluyen con la elaboración de un instrumento, denominado “Rubric for the Evaluation of Educational App for preschool Children (REVEAC)”, para evaluar la calidad de App educativas dirigidas a niños en edad preescolar. En ella, se consideran dimensiones como el contenido educativo, diseño, funcionalidad y características técnicas. En la parte del contenido educativo se tuvieron en cuenta aspectos como la edad y nivel de aprendizaje de los usuarios, su pensamiento crítico, el aprendizaje múltiple y la evaluación.

La consideración del desarrollo de Apps con finalidad educativa y su proceso de evaluación debe considerar aspectos relevantes en los procesos de aprendizaje en la sociedad actual. En los modelos constructivistas el sujeto no solo consume información, sino que desarrolla la manipulación y construcción de forma previa al consumo de ésta.

Es decir, se aprende por la experiencia y manipulación, como ya planteaba Piaget. Si además se incorporan a dicho razonamiento las consideraciones propias del constructivismo social de Vygostky (Duque & Packer, 2014), que postulan que las funciones superiores de pensamiento y que los aprendizajes parten del ámbito social, para posteriormente incorporarse individualmente, se puede constatar la importancia y relevancia de los nuevos escenarios digitales como espacios actuales de interacción social y manipulación de la información generada, así como elementos básicos del aprendizaje del individuo digital.

Si bien existen propuestas de evaluación de aplicaciones para su utilización en contextos educativos, no se han encontrado protocolos que planteen la evaluación desde una perspectiva multidimensional y centrados en analizar si una App puede contribuir al desarrollo del individuo desde diferentes planos: la personalidad, la inteligencia emocional, los procesos cognitivos y metacognitivos, el aprendizaje y desarrollo de competencias, así como otros aspectos vinculados a la integración social, sin dejar de lado aspectos de carácter técnico como la usabilidad, la accesibilidad o el soporte al usuario.

Esta aportación plantea el diseño y validación de contenido de una rúbrica para evaluar la idoneidad de las Apps para escenarios de enseñanza-aprendizaje. La rúbrica, como instrumento de registro observacional, puede favorecer los procesos de evaluación. Una rúbrica permite establecer una estructura ordenada y dimensional de criterios de evaluación, facilitar una forma de obtener una puntuación, tanto global como por dimensión, y simplifica la selección del nivel de cumplimiento en cada criterio de evaluación observable mediante un descriptor explicativo en cada nivel (Oakleaf, 2009).

La generación y validación de una rúbrica por medio de procedimientos científicamente aceptables, como se describen en el siguiente epígrafe, resulta por tanto relevante desde el punto de vista educativo. En este sentido, los objetivos planteados para este estudio son los siguientes: 1) Delimitar estos elementos educativos y formativos, de una aplicación o App, como variables observables y/o mensurables mediante un instrumento ad hoc; 2) Elaborar un protocolo de comprobación del valor lúdico-educativo y formativo de aplicaciones o App; y 3) Especificar los elementos o criterios del valor educativo de una App y el protocolo de comprobación, mediante la construcción de una rúbrica de valoración del valor educativo de Apps.

2. Método

El método para el diseño, construcción y validación de la rúbrica se ajusta a los procedimientos generales de construcción y a los criterios específicos para rúbricas evaluativas de Apps (Cherner et al., 2016). Para ello, se plantean una serie de fases con el desarrollo de una metodología mixta que contempla procedimientos cualitativos y cuantitativos: Fase 1) Diseño conceptual y teórico de la rúbrica; Fase 2) Diseño funcional de la rúbrica a nivel de dimensiones, indicadores y descriptores de valoración de cada indicador; y Fase 3) Validación de contenido teórico-práctica, mediante juicio de expertos en cuanto a la pertinencia, adecuación y claridad, siguiendo la técnica Delphi x 3.

La primera fase de diseño conceptual y teórico de la rúbrica se genera desde las apreciaciones reflexivas realizadas por el grupo de investigación especializado en el campo de la educación y de recursos digitales, que plantea un proceso de revisión de la literatura desde una perspectiva conceptual y metodológica. Es decir, la búsqueda, recuperación y revisión de trabajos sobre evaluación de Apps que faciliten la generación de dimensiones e indicadores sobre su valor educativo y sobre los procedimientos metodológicos fundamentados para la construcción de la propia rúbrica.

Una vez establecidas las dimensiones y sub-dimensiones consideradas desde el ámbito teórico, la segunda fase se centró en el proceso de búsqueda de indicadores que nos permitan describir en profundidad cada una de las dimensiones, dando lugar a un inventario de ítems.

2.1. Diseño conceptual y teórico de la rúbrica

Las primeras fases se realizan desde una orientación teórica, tanto normativa como epistemológica. Para decidir los aspectos evaluativos de carácter técnico se utilizó la Directiva Europea UE 2016/2102 (European Parliament, 2016), así como en la normativa española (Real Decreto 1112/2018, de 7 de septiembre, sobre accesibilidad de los sitios web y aplicaciones para dispositivos móviles del sector público, 2018), que como ya se ha descrito se limita al planteamiento de la accesibilidad para facilitar su uso a todas las personas. También se han consultado informes realizados por autoridades educativas y fundaciones o asociaciones con finalidad educativa, que se centran más en estadísticas de uso y aspectos tecnológicos, que en su uso o potencialidad educativa (European Commission, 2021).

La revisión de la literatura científica recomienda, además del uso de variables o dimensiones tecnológicas, la incorporación de aspectos educativos como elementos para enjuiciar la calidad y el potencial educativo de las Apps. En este sentido, se realiza una síntesis integrativa (ver Tabla 1) en la que se puede constatar el ajuste de los indicadores resultantes a los tópicos educativos planteados por la UNESCO (UNESCO, 2019), las estrategias cognitivas y metacognitivas planteadas desde la teoría de las inteligencias múltiples de Gardner, dado su interés de cara a promover en el aula la creatividad, la maduración, el desempeño escolar y los procesos de comunicación (Abenti, 2020; Berrios-Aguayo et al., 2021; Kalimaya et al., 2021). Igualmente, se han tenido en cuenta los procesos y actividades cognitivas basados en las TIC, siguiendo la taxonomía de Bloom en su definición más actual (Krathwohl & Anderson, 2010), el modelo SAMR (Substitution, Augmentation, Modification, Redefinition) y el modelo TPACK (Technological Pedagogical Content Knowledge), que es un modelo basado en la idea de combinar tres variables, el conocimiento tecnológico, el conocimiento pedagógico y el conocimiento del contenido (Puentedura, 2014).

Por otra parte, en la rúbrica, reflejada como resultado o producto final, se plasman los campos de trabajo para el desarrollo de los procesos educativos y del aprendizaje que plantean algunos autores (Expósito-López, 2018; Expósito-López & Olmedo-Moreno, 2020; Harrison, 2009). Por otro lado, partiendo del modelo SAMR y las consideraciones de Puentedura (2014) para mejorar el modelo. En base a este proceso se plantean las

variables, dimensiones e indicadores para valorar educativa y tecnológicamente las Apps que se muestra en la Tabla 1.

Tabla 1.

Resultados de indicadores mensurables planteados según valor educativo y tecnológico

VARIABLES	DIMENSIONES	INDICADORES
Variables educativas intrínsecas	Mejora o ayuda a desarrollar la personalidad del usuario	Autoestima y auto-concepto Asertividad Inteligencia emocional Responsabilidad y tesón
	Mejora del proceso de aprendizaje del usuario	Motivación Integración de nuevos conocimientos y contenidos Estrategias cognitivas Estrategias meta-cognitivas Técnicas de estudio/aprendizaje
	Mejora de la interacción social del usuario	Empatía Pensamiento crítico y apertura mental Respeto a las normas y valores sociales Participación activa Sociabilidad y extroversión
Variables educativas asociadas al soporte tecnológico	Mejora del acceso del usuario-máquina	Accesibilidad Interacción Retroalimentación Adaptabilidad
	Mejora de la interacción usuario-máquina como proceso de aprendizaje	Estructura Navegación Versatilidad Usabilidad
	Mejora del feedback educativo	Recompensa Integralidad del conocimiento provisto Finalidad estratégica como proceso formativo

Fuente: Elaboración propia.

En el procedimiento de revisión preliminar participaron tres expertos, realizando sugerencias de modificación en las dimensiones y descriptores de la rúbrica la cual quedó integrada por un total de 21 bloques de indicadores de evaluación, agrupados en un conjunto de 5 dimensiones tal y como se muestra en la Tabla 2.

Tabla 2.

Resultados preliminares de dimensiones y bloques de ítems para evaluar el valor educativo e las App

DIMENSIONES	BLOQUES
Dimensión de desarrollo personal	1. Desarrollo de aspectos referidos a la personalidad del usuario. 2. Desarrollo de la inteligencia emocional: Conciencia emocional 3. Desarrollo de la inteligencia emocional: Regulación emocional 4. Desarrollo de la inteligencia emocional: Autonomía emocional

	5. Desarrollo de la inteligencia emocional: Competencias para la vida
	6. Desarrollo de la inteligencia emocional: Competencias para el bienestar
Dimensión de desarrollo cognitivo y metacognitivo	7. Empleo y desarrollo de estrategias cognitivas de ensayo
	8. Empleo y desarrollo de estrategias cognitivas de elaboración
	9. Empleo y desarrollo de estrategias cognitivas organizativas
	10. Empleo y desarrollo de estrategias metacognitivas sobre la tarea
	11. Empleo y desarrollo de estrategias metacognitivas sobre el proceso de aprendizaje
	12. Empleo y desarrollo de estrategias metacognitivas de apoyo
Dimensión de desarrollo del aprendizaje y de competencias	13. Fomento del proceso general de aprendizaje del usuario
	14. Desarrollo de competencias comunicativas
	15. Desarrollo de competencias artísticas y culturales
	16. Desarrollo de competencias lógico-matemáticas
	17. Desarrollo de competencias físico- kinestésicas
Dimensión del desarrollo social	18. Desarrollo de aspectos referidos a las habilidades sociales y a la integración social
Dimensión técnica	19. Usabilidad
	20. Accesibilidad
	21. Compatibilidad

Fuente: Elaboración propia.

2.2. Selección de jueces expertos

Este inventario se sometió a revisión empleando un proceso Delphi con tres vueltas, en el que participaron 14 jueces evaluadores expertos que fueron seleccionados en base a los siguientes criterios tal y como establece Arteaga-Martínez et al. (2021): compromiso y participación en todo el proceso, experiencia académica e investigadora en el objeto de estudio, pluralidad de planteamientos en los procesos de aprendizaje y uso de recursos digitales en dichos procesos, conocimiento de la técnica y/o metodología de revisión.

Esta aproximación puede ser descrita como un método de estructuración de un proceso de comunicación grupal que es efectivo a la hora de permitir a un grupo de individuos tratar un problema complejo. Lo que se persigue es obtener el grado de consenso o acuerdo entre especialistas sobre el problema planteado (Arteaga-Martínez et al., 2021), que en este caso se especifica en el acuerdo sobre el inventario de dimensiones e ítems sometidos a validación.

Inicialmente, se llevó a cabo una revisión preliminar para mejorar el instrumento, en la que participaron tres expertos a los que se planteó un conjunto inicial de dimensiones, indicadores y descriptores, los cuales tras un proceso de valoración cualitativa realizaron sugerencias de modificación. Posteriormente, se convocó a través de correo electrónico a un panel de 35 personas expertas, de las cuales 14 aceptaron colaborar con la investigación, con los siguientes perfiles profesionales: 3 eran expertos académicos en el ámbito de la comunicación e informática, 6 especialistas en metodologías de la

investigación educativa, 2 profesionales en el sector de las nuevas tecnologías y 3 expertos en desarrollo de aplicaciones para dispositivos móviles (ver Tabla 3).

En la Tabla 3 se puede ver que estas personas obtuvieron un Coeficiente de Competencia Experta K superior al nivel crítico exigible de 0,8 calculado a partir del Coeficiente de Conocimiento Kc y el Coeficiente de Argumentación Ka (Cabero-Almenara & Llorente-Cejudo, 2013). Como fuentes de argumentación los expertos indicaron la experiencia obtenida a través de su actividad práctica y profesional (Exp), su nivel de conocimiento acerca del problema de investigación (Con), su intuición sobre el tema (Int), y los estudios realizados en torno a la temática (Est), cuya sumatoria permite obtener el Ka (Cabero-Almenara & Barroso-Osuna, 2013).

Tabla 3.
Coeficiente de Competencia Experta para la selección del Panel de Expertos

Cód	Perfil Profesional	Estudios	Exp	Con	Int	Est	Ka	Kc	K
E1	Académico - Comunicación, informática	Doctorado	,5	,3	,05	,03	,88	,9	,88
E2	Académico - Comunicación e informática	Doctorado	,5	,3	,05	,05	,9	,79	,84
E3	Académico - Comunicación e informática	Doctorado	,5	,2	,05	,05	,8	,86	,83
E4	Académico - Metodólogo	Doctorado	,5	,3	,1	,05	,95	,8	,87
E5	Académico - Metodólogo	Doctorado	,5	,3	,1	,05	,95	,8	,87
E6	Académico - Metodólogo	Doctorado	,5	,3	,1	,03	,93	,78	,85
E7	Académico - Metodólogo	Doctorado	,5	,3	,1	,03	,93	,78	,85
E8	Académico - Metodólogo	Máster	,5	,3	,05	,03	,88	,9	,89
E9	Académico - Metodólogo	Máster	,4	,3	,05	,05	,8	,9	,85
E10	Consultor nuevas tecnologías	Doctorado	,5	,3	,1	,05	,95	,8	,87
E11	Consultor nuevas tecnologías	Máster	,5	,3	,05	,05	,9	,8	,85
E12	Desarrollador software	Ingeniería	,5	,3	,05	,03	,88	,8	,84
E13	Desarrollador software	Ingeniería	,5	,3	,05	,03	,88	,79	,83
E14	Desarrollador software	Ingeniería	,5	,3	,1	,03	,93	,78	,85

Fuente: Elaboración propia.

Los jueces expertos efectuaron una valoración cuantitativa de cada uno de los ítems de la primera versión de la rúbrica según criterios de pertinencia, adecuación y claridad, empleando una escala de tipo Likert de 5 opciones de respuesta, basándonos en la propuesta de Salazar-Gómez et al. (2018), e incorporando una dimensión de evaluación adicional para valorar la claridad de los ítems. El procedimiento se llevó a cabo hasta lograr depurar los juicios de grupo mediante un procedimiento matemático de agregación de juicios individuales. La Tabla 4 refleja los criterios utilizados por el equipo de jueces expertos

Tabla 4.

Elementos empleados en el juicio de expertos para la evaluación de la rúbrica

Categoría	Clasificación	Criterio
Pertinencia El ítem permite evaluar un aspecto central del propósito, constructo teórico y/o dimensiones del instrumento	(1) No es pertinente	El ítem no contribuye a evaluar el propósito o dimensiones del instrumento en ningún aspecto. Puede ser eliminado.
	(2) Bajo nivel de pertenencia	El ítem hace una contribución parcial a la evaluación del propósito o dimensiones del instrumento. De ser relaborado.
	(3) Aceptable nivel de pertenencia	El ítem contribuye a la evaluación del propósito o dimensiones del instrumento. Puede mejorar se su contenido
	(4) Bastante nivel de pertinencia	El ítem contribuye bastante al grado de evaluación del propósito o dimensiones del instrumento.
	(5) Alto nivel de pertenencia	El ítem contribuye en un alta grado a la evaluación del propósito o dimensiones del instrumento.
Adecuación El ítem se ajusta a la dimensión evaluativa y aspectos que la componen	(1) No es adecuado	El ítem no se ajusta a la dimensión evaluativa o dimensión del instrumento en ningún aspecto. Puede ser eliminado.
	(2) Bajo nivel de adecuación	El ítem posee un ajuste parcial a la evaluación del propósito o dimensión del instrumento. De ser relaborado o recolocado.
	(3) Aceptable nivel de adecuación	El ítem se ajusta a la evaluación del propósito o dimensión del instrumento. Puede mejorar se su contenido.
	(4) Bastante nivel de adecuación	El ítem se ajusta bastante a la evaluación del propósito o dimensión del instrumento.
	(5) Alto nivel de adecuación	El ítem se ajusta en un alta grado a la evaluación del propósito o dimensión del instrumento.
Claridad El ítem es comprensible por los potenciales	(1) No es claro	El ítem no es comprensible por los potenciales usuarios del instrumento. Debe ser eliminado
	(2) Bajo nivel de claridad	El ítem puede ser mejorado al menos en la mitad de sus componentes, en aspectos de redacción y gramática.

usuarios y se adecua a las reglas gramaticales	(3) Aceptable nivel de claridad	El ítem puede mejorarse en algún elemento para mejorar su comprensión
	(4) Bastante nivel de claridad	El ítem es bastante claro, pero puede mejorar su comprensión
	(5) Alto nivel de claridad	El ítem es comprendido altamente por los potenciales usuarios y cumple las reglas gramaticales.

Fuente: Salazar-Gómez et al. (2018)

Para facilitar el proceso de evaluación, se procedió a la digitalización de la escala propuesta utilizando un formulario online de Google, que se hizo llegar al panel de expertos. El formulario fue estructurado en torno al conjunto total de dimensiones e ítems, incluyendo campos numéricos para seleccionar el nivel de pertinencia, adecuación y claridad, así como un campo opcional para realizar valoraciones de carácter cualitativo.

3. Análisis de datos

El proceso de análisis de datos responde según la naturaleza de los datos considerados, a la información recopilada en el proceso Delphi, que supone la validación de contenidos de la rúbrica, así como a los datos de carácter cuantitativo obtenidos tras el procedimiento de juicio de expertos. Tras la evaluación realizada por el panel de expertos (n=14), se recogieron los datos del formulario online y se llevó a cabo un proceso de análisis de concordancia entre evaluadores, desde dos planos. Inicialmente, se estudió el nivel de acuerdo global mediante el cálculo del Coeficiente de Correlación Intraclase (CCI) tanto para el total de ítems de la rúbrica como para cada una de las agrupaciones de ítems en las dimensiones en las que se estructuró. Este coeficiente refleja la variación entre varios evaluadores que miden el mismo grupo de ítems (Koo & Li, 2016). Para su cálculo se utilizó el paquete irr 0.84.1 en la versión 4.1.1 del software estadístico R, considerando un nivel de confianza del 95%.

Posteriormente, los datos fueron trasladados a una hoja de cálculo construida ad-hoc, en la que se calculó, para cada ítem, la media de las evaluaciones y el coeficiente de validez de contenido V de Aiken en las tres dimensiones de pertinencia, adecuación y claridad. Para el cálculo se siguió la modificación propuesta por Penfield & Giacobbi (2004), considerando la media de las valoraciones de los evaluadores, el valor mínimo de la escala ($l=1$) y la diferencia entre los valores máximo y mínimo de la escala ($k=4$).

$$V = \frac{\bar{X} - l}{k}$$

El valor del estadístico V de Aiken puede oscilar entre 0 que se interpreta como ausencia total de acuerdo y 1 que supone un nivel de acuerdo óptimo, por tanto, aquellos valores del estadístico cercanos a 1 evidencian una validez de contenido. Inicialmente, en las dimensiones de pertinencia y adecuación, siguiendo la recomendación de Charter (2003), se descartaron aquellos ítems con una V de Aiken inferior a 0,7.

Además, se ha calculó en cada caso el intervalo de confianza (IC), empleando el método Score gracias a su independencia sobre la normalidad de la variable, la asimetría respecto a dicha variable y su exactitud (Merino-Soto & Livia-Segovia, 2009), considerando el valor de z a 1,96 (distribución estándar al 95%), obteniendo los límites inferior y superior utilizando los siguientes cálculos:

$$IC\ LI = \frac{2nkV + z^2 - z\sqrt{4nkV(1 - V) + z^2}}{k}$$

$$IC\ LS = \frac{2nkV + z^2 + z\sqrt{4nkV(1 - V) + z^2}}{k}$$

En un segundo filtro, se tuvieron solo en cuenta aquellos ítems cuyo límite inferior del IC no alcanzó el umbral de 0,7 (Penfield & Giacobbi, 2004). Tras la iteración final del proceso Delphi, se eliminaron de la versión final de la rúbrica aquellos indicadores en los que el nivel de acuerdo según el límite inferior se mantuvo bajo el umbral de 0,7 en las dimensiones de pertinencia o adecuación. Por otro lado, aquellos ítems inicialmente aceptados según su pertinencia y adecuación, pero con índices bajos en la dimensión de claridad fueron reformulados de cara a la siguiente ronda de revisión y a la versión final.

4. Resultados de la validación

4.1. Resultados globales del acuerdo entre jueces expertos

La revisión definitiva con la participación de 14 jueces expertos produce resultados que se obtienen al calcular el Coeficiente de Correlación Intraclase (ver Tabla 5) denotan un nivel de acuerdo global bajo en cada dimensión ($CCI < 0,5$), siendo ligeramente superior en las dimensiones de desarrollo social y técnica. No obstante, es importante destacar que en todos los casos analizados el resultado es estadísticamente significativo ($p < 0,05$) según la Prueba F de Fisher-Snedecor, lo cual implica que los resultados obtenidos no se deben al azar.

Tabla 5.
Coefficientes de Correlación Intraclase por dimensiones

	CCI	IC (95%)		Prueba F			Sig
		LI	LS	Valor	gl1	gl2	
Desarrollo personal	,078	,018	,201	2,182	21	273	,002
Desarrollo cognitivo y metacognitivo	,144	,064	,294	3,364	22	286	,000
Desarrollo del aprendizaje y competencias	,109	,038	,248	2,705	21	273	,000
Desarrollo social	,254	,096	,595	5,778	8	104	,000
Dimensión técnica	,214	,138	,327	4,822	44	572	,000

Fuente: Elaboración propia.

Posteriormente, en un enfoque centrado en cada dimensión / ítem, se llevó a cabo el cálculo del coeficiente de validez V de Aiken junto a los límites inferior (LI) y superior (LS), considerando un nivel de confianza del 95% y habiendo llevado a cabo tres rondas de revisión. Esta aproximación ha facilitado la toma de decisiones en torno a los ítems que aceptar tal y como se sometieron a evaluación, aquellos que reformular modificando la redacción original, o bien eliminar de la versión final de la rúbrica. En las tablas que exponen el resultado de la ejecución del coeficiente, se indica la decisión adoptada en cada ítem: O) Mantener original, T) Indicador transformado, E) Indicador eliminado.

La Tabla 6 resume el resultado del proceso de validación de contenido tras aplicar la técnica Delphi en tres rondas. De los 121 ítems que se sometieron inicialmente a juicio de expertos, 21 fueron eliminados y 15 fueron reformulados por haberse considerado pertinentes y adecuados al constructo, pero no suficientemente claros en su redacción.

Tabla 6.
Resultados del proceso de validación de contenido

Dimensión evaluativa	Bloques de ítems	Ítems sometidos	Ítems reformulados	Ítems eliminados
Desarrollo personal	6	22	1	4
Desarrollo cognitivo y metacognitivo	6	23	2	5
Desarrollo del aprendizaje y competencias	5	22	3	5
Desarrollo social	1	9	2	2
Aspectos técnicos	3	45	7	5
	21	121	15	21

Fuente: Elaboración propia.

A continuación, se describen los resultados de la valoración en base a estas dimensiones.

4.2. Resultados de la dimensión de desarrollo personal

Esta primera dimensión de la rúbrica evalúa la capacidad de las Apps para desarrollar determinados aspectos de la personalidad, así como la inteligencia emocional de los potenciales usuarios. La Tabla 7 muestra los resultados obtenidos tras la ronda final de validación en los seis primeros bloques de indicadores. De los ítems 1.1 a 1.5, que miden si una App ayuda al usuario a desarrollar aspectos referidos a la personalidad del usuario, se decidió eliminar 2 de ellos, tras arrojar valores inferiores al umbral de 0,7 tanto en pertinencia como en adecuación, quedando la dimensión compuesta por un total de 3 indicadores (ver apartado de material suplementario).

Tabla 7.

Cálculos V de Aiken e IC sobre bloques del desarrollo personal (bloques 1 a 6)

Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec
P	4,50	,88	,76	,94		P	4,07	,77	,64*	,86		3.1 P	4,43	,86	,74	,93	
1.1 A	4,50	,88	,76	,94	O	2.1 A	3,86	,71	,59*	,82	E	A	4,50	,88	,76	,94	O
C	4,36	,84	,72	,91		C	4,29	,82	,70	,90		C	4,50	,88	,76	,94	
P	4,79	,95	,85	,98		P	4,43	,86	,74	,93		3.2 P	4,71	,93	,83	,97	
1.2 A	4,71	,93	,83	,97	O	2.2 A	4,29	,82	,70	,90	O	A	4,50	,88	,76	,94	O
C	4,43	,86	,74	,93		C	4,57	,89	,79	,95		C	4,36	,84	,72	,91	
P	4,57	,89	,79	,95		P	4,50	,88	,76	,94		3.3 P	4,50	,88	,76	,94	
1.3 A	4,29	,82	,70	,90	O	2.3 A	4,50	,88	,76	,94	O	A	4,50	,88	,76	,94	O
C	4,50	,88	,76	,94		C	4,64	,91	,81	,96		C	4,50	,88	,76	,94	
P	4,07	,77	,64*	,86		P	4,71	,93	,83	,97							
1.4 A	4,07	,77	,64*	,86	E	2.4 A	4,43	,86	,74	,93	O						
C	4,29	,82	,70	,90		C	4,43	,86	,74	,93							
P	4,00	,75	,62*	,84													
1.5 A	3,86	,71	,59*	,82	E												
C	4,07	,77	,64*	,86													
P	4,57	,89	,79	,95		P	4,36	,84	,72	,91		P	4,50	,88	,76	,94	
4.1 A	4,29	,82	,70	,90	O	5.1 A	4,50	,88	,76	,94	O	6.1 A	4,29	,82	,70	,90	O
C	4,36	,84	,72	,91		C	4,29	,82	,70	,90		C	3,86	,71	,59*	,82	
P	4,50	,88	,76	,94		P	4,50	,88	,76	,94		P	4,71	,93	,83	,97	
4.2 A	4,29	,82	,70	,90	T	5.2 A	4,57	,89	,79	,95	O	6.2 A	4,36	,84	,72	,91	O
C	4,43	,86	,74	,93		C	4,50	,88	,76	,94		C	4,93	,98	,91	1,00	
P	4,43	,86	,74	,93		P	4,50	,88	,76	,94		P	4,36	,84	,72	,91	
4.3 A	4,36	,84	,72	,91	T	5.3 A	4,64	,91	,81	,96	O	6.3 A	4,50	,88	,76	,94	O
C	4,36	,84	,72	,91		C	4,50	,88	,76	,94		C	4,43	,86	,74	,93	
						P	4,00	,75	,62*	,84							
						5.4 A	3,71	,68*	,55*	,79	E						
						C	4,29	,82	,70	,90							

Fuente: Elaboración propia.

P: Pertinencia, A: Adecuación, C: Claridad

IC LI: Int. de confianza límite inferior, IC LS: Int. de confianza límite superior

O: Mantener original, T: Indicador transformado, E: Indicador eliminado

* Valores fuera de los límites admisibles

En la valoración de los ítems relativos a medir si el uso de una App puede contribuir al desarrollo de la conciencia emocional (2.1 a 2.4), se procedió a eliminar un ítem que tras la última vuelta de revisión obtuvo valores del límite inferior del coeficiente V de Aiken en pertinencia y adecuación de ,64 y ,59, quedando finalmente este bloque compuesto por 3 ítems. En lo que respecta a la capacidad de las Apps para desarrollar la competencia de regulación emocional (bloque 3), no fue preciso modificar o eliminar ninguno de los ítems inicialmente sometidos a evaluación, situándose todos los valores del estadístico por encima de ,8.

El bloque 4 se encarga de medir la capacidad de una App para promover la autonomía emocional y a tenor de los resultados obtenidos no fue necesario alterar la propuesta inicial de 3 indicadores, dado que se obtuvieron valores aceptables en los coeficientes y valores del límite inferior, sin embargo, se optó por reformular dos ítems por haberse recomendado una mejor redacción en las valoraciones cualitativas vertidas aun habiendo superado el umbral exigido en la dimensión de claridad.

En el bloque 5, que mide si la App facilita el desarrollo de competencias para la vida, se encontraron valores adecuados en el coeficiente V de Aiken en todos los indicadores de evaluación salvo en uno de ellos, que tras las sucesivas reformulaciones en las vueltas del proceso Delphi, no logró alcanzar un nivel suficiente de pertinencia según el límite inferior, ni de adecuación según el coeficiente. En lo relativo a la posibilidad de una App para desarrollar competencias para el bienestar (bloque de indicadores 6), se puede observar que se mantuvieron todos ítems planteados, aunque se reformuló el primero de ellos tras obtener un valor de ,59 en el LI de la dimensión de claridad.

4.3. Resultados de la dimensión de desarrollo cognitivo y metacognitivo

En esta dimensión, el bloque 7 permite medir si una app requiere que el usuario emplee y desarrolle estrategias cognitivas de ensayo. En la Tabla 8 puede observarse que, de los 4 indicadores inicialmente sometidos, se decidió eliminar uno de ellos por haber encontrado valores bajos en pertinencia (LI=,64) y adecuación (V=,68). Por otra parte, uno de los ítems fue reformulado tras detectarse poca claridad (LI=,59).

Tabla 8.
 Cálculos V de Aiken e IC - bloques 7, 8, 9, 10 y 11

Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec
P	4,43	,86	,74	,93		P	4,43	,86	,74	,93		P	4,71	,93	,83	,97	
7.1 A	4,71	,93	,83	,97	T	8.1 A	4,29	,82	,70	,90	O	9.1A	4,64	,91	,81	,96	O
C	3,86	,71	,59*	,82		C	4,36	,84	,72	,91		C	4,71	,93	,83	,97	
P	4,07	,77	,64*	,86		P	4,50	,88	,76	,94							
7.2 A	3,71	,68*	,55*	,79	E	8.2 A	4,36	,84	,72	,91	O						
C	4,71	,93	,83	,97		C	4,64	,91	,81	,96							
P	4,43	,86	,74	,93		P	4,50	,88	,76	,94							
7.3 A	4,43	,86	,74	,93	O	8.3 A	4,50	,88	,76	,94	O						
C	4,50	,88	,76	,94		C	4,43	,86	,74	,93							
P	4,64	,91	,81	,96													
7.4 A	4,64	,91	,81	,96	O												
C	4,29	,82	,70	,90													
P	4,50	,88	,76	,94		P	4,57	,89	,79	,95		P	4,57	,89	,79	,95	
10.1A	4,57	,89	,79	,95	O	11.1A	4,43	,86	,74	,93	O	12.1A	4,50	,88	,76	,94	O
C	4,50	,88	,76	,94		C	4,29	,82	,70	,90		C	4,36	,84	,72	,91	
P	4,50	,88	,76	,94		P	4,36	,84	,72	,91		P	3,71	,68*	,55*	,79	
10.2A	4,86	,96	,88	,99	O.	11.2A	4,50	,88	,76	,94	O	12.2A	3,93	,73	,60*	,83	E
C	4,36	,84	,72	,91		C	4,43	,86	,74	,93		C	4,36	,84	,72	,91	

P 4,43,86 ,74 ,93	P 4,07,77 *,64 ,86	P 4,64 ,91 ,81 ,96
10.3A 4,64,91 ,81 ,96 T	11.3A 4,00,75 *,62 ,84 E	12.3A 4,71 ,93 ,83 ,97 O
C 3,79,70 ,57*,80	C 4,50,88 ,76 ,94	C 4,71 ,93 ,83 ,97
P 4,71,93 ,83 ,97	P 4,71,93 ,83 ,97	P 4,43 ,86 ,74 ,93
10.4A 4,36,84 ,72 ,91 O	11.4A 4,57,89 ,79 ,95 O	12.4A 4,50 ,88 ,76 ,94 O
C 4,50,88 ,76 ,94	C 4,43,86 ,74 ,93	C 4,50 ,88 ,76 ,94
	P 3,79,70 ,57*,80	P 4,36 ,84 ,72 ,91
	11.5A 3,64,66*,53*,77 E	12.5A 4,64 ,91 ,81 ,96 O
	C 4,29,82 ,70 ,90	C 4,57 ,89 ,79 ,95
		P 3,86 ,71 ,59*,82
		12.6A 3,86 ,71 ,59*,82 E
		C 4,00 ,75 ,62*,84

Fuente: Elaboración propia.

P: Pertinencia, A: Adecuación, C: Claridad

IC LI: Int. de confianza límite inferior, IC LS: Int. de confianza límite superior

O: Mantener original, T: Indicador transformado, E: Indicador eliminado

* Valores fuera de los límites admisibles

En las dimensiones de desarrollo de estrategias cognitivas de elaboración y desarrollo de estrategias cognitivas organizativas (bloques 8 y 9) no se eliminó ninguno de los ítems propuestos en base a los criterios de pertinencia y adecuación. Tampoco fue preciso reformular ningún indicador en base a los resultados obtenidos. Como puede verse en la Tabla 8, se reformuló un ítem del bloque 10, destinado a medir la posibilidad de una App para desarrollar estrategias metacognitivas sobre la tarea (LI=,57 en claridad). Por otro lado, se tomó la decisión de eliminar dos ítems correspondientes inicialmente al bloque 11 (estrategias metacognitivas sobre el proceso de aprendizaje). En concreto el 11.3, LI=,64 en pertinencia y LI=,62 en adecuación, y también el 11.5 habiéndose obtenido un LI de ,57 en pertinencia y no alcanzándose el umbral requerido en la VA en adecuación tras las sucesivas iteraciones del proceso Delphi. En la dimensión sobre estrategias metacognitivas de apoyo (bloque 12), puede observarse que dos indicadores fueron eliminados a tenor de los valores obtenidos (12.2 y 12.6).

4.4. Resultados de la dimensión sobre desarrollo del aprendizaje y de competencias

En el primer bloque de ítems de esta dimensión, que evalúa si el uso de una App fomenta el proceso general de aprendizaje del usuario (bloque 13), fue necesario reformular dos ítems que tras las iteraciones mantuvieron valores inferiores al umbral en el LI de claridad: ,59 y ,68 respectivamente (ver Tabla 9). En los indicadores correspondientes a medir si una App facilita el desarrollo de la competencia comunicativa (bloque 14), se eliminó un ítem al mantener tras la última iteración valores bajos del LI tanto en pertinencia (,60) como en adecuación (,66) si bien el valor de V superaba el umbral. Por otro lado, se puede observar que, de los 5 ítems que se sometieron a evaluación en el bloque 15, que mide si una App fomenta competencias artísticas y culturales en los usuarios, solo tres se mantuvieron en la versión final. El indicador 15.2 presentó un nivel de acuerdo de los evaluadores no suficiente en torno a su pertinencia (V=,7; LI=,57) y adecuación (LI=,60), por lo que se decidió eliminarlo de la versión final. El ítem 15.5 también debió eliminarse al no alcanzarse valores

aceptables tanto en pertinencia como en adecuación. El bloque quedó compuesto por tres indicadores tras reformular uno de ellos al presentar un valor de ,59 en el LI de claridad. Respecto a la capacidad de una app para fomentar competencias lógico-matemáticas (bloque 16), fue necesario eliminar uno de los ítems que se sometió a evaluación, no habiendo detectado suficiente nivel de acuerdo entre los evaluadores en torno a su pertinencia (LI=,64) y su adecuación (LI=,62).

Tabla 9.
 Cálculos V de Aiken e IC - bloques 13, 14, 15, 16 y 17

Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec
P	4,36	,84	,72	,91		P	4,57	,89	,79	,95		P	4,50	,88	,76	,94	
13.1 A	4,29	,82	,70	,90	O	14.1 A	4,50	,88	,76	,94	O	15.1 A	4,64	,91	,81	,96	T
C	4,71	,93	,83	,97		C	4,36	,84	,72	,91		C	3,86	,71	,59*	,82	
P	4,29	,82	,70	,90		P	4,50	,88	,76	,94		P	3,79	,70	,57*	,80	
13.2 A	4,50	,88	,76	,94	T	14.2 A	4,29	,82	,70	,90	O	15.2 A	3,93	,73	,60*	,83	E
C	3,86	,71	,59*	,82		C	4,36	,84	,72	,91		C	4,36	,84	,72	,91	
P	4,57	,89	,79	,95		P	3,93	,73	,60*	,83		P	4,29	,82	,70	,90	
13.3 A	4,64	,91	,81	,96	T	14.3 A	4,14	,79	,66*	,87	E	15.3 A	4,29	,82	,70	,90	O
C	4,21	,80	,68*	,89		C	4,50	,88	,76	,94		C	4,57	,89	,79	,95	
						P	4,57	,89	,79	,95		P	4,43	,86	,74	,93	
						14.4 A	4,64	,91	,81	,96	O	15.4 A	4,43	,86	,74	,93	O
						C	4,50	,88	,76	,94		C	4,57	,89	,79	,95	
						P	4,43	,86	,74	,93		P	4,14	,79	,66*	,87	
						14.5 A	4,57	,89	,79	,95	O	15.5 A	3,71	,68*	,55*	,79	E
						C	4,57	,89	,79	,95		C	4,29	,82	,70	,90	

Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec
P	4,57	,89	,79	,95		P	4,64	,91	,81	,96	
16.1 A	4,64	,91	,81	,96	O	17.1 A	4,50	,88	,76	,94	O
C	4,29	,82	,70	,90		C	4,43	,86	,74	,93	
P	4,07	,77	,64*	,86		P	4,43	,86	,74	,93	
16.2 A	4,00	,75	,62*	,84	E	17.2 A	4,43	,86	,74	,93	O
C	3,71	,68*	,55*	,79		C	4,57	,89	,79	,95	
P	4,71	,93	,83	,97		P	4,50	,88	,76	,94	
16.3 A	4,50	,88	,76	,94	O	17.3 A	4,43	,86	,74	,93	O
C	4,50	,88	,76	,94		C	4,50	,88	,76	,94	
P	4,43	,86	,74	,93		P	4,14	,79	,66*	,87	
16.4 A	4,57	,89	,79	,95	O	17.4 A	3,93	,73	,60*	,83	E
C	4,29	,82	,70	,90		C	4,07	,77	,64*	,86	
						P	4,79	,95	,85	,98	
						17.5 A	4,57	,89	,79	,95	T
						C	3,79	,70	,57*	,80	

P: Pertinencia, A: Adecuación, C: Claridad

IC LI: Int. de confianza límite inferior, IC LS: Int. de confianza límite superior

O: Mantener original, T: Indicador transformado, E: Indicador eliminado

* Valores fuera de los límites admisibles

Finalmente, en el bloque 17, que mide si una app puede fomentar las competencias físico-kinestésicas, se mantuvieron 4 ítems y se eliminó 1 por presentar valores bajos de acuerdo tanto en pertinencia (LI=,66), como en adecuación (LI=,60). También fue necesario reformular el indicador 17.5 por baja claridad (LI=,57).

4.5. Resultados de la dimensión de desarrollo social

Esta dimensión está compuesta por un único bloque (18) que incluye un conjunto de ítems para medir si una App puede fomentar el desarrollo de aspectos referidos a las habilidades sociales y la integración social de los usuarios. Inicialmente se sometieron a evaluación 9 ítems y tras los cálculos del coeficiente y los límites se decidió eliminar 2 ítems y reformular otros 2.

Como puede verse en la Tabla 10, los ítems eliminados presentaron niveles de acuerdo bajos en todos los aspectos analizados (pertinencia, adecuación y claridad). En uno de los ítems reformulados se obtuvo ,63 en la V de Aiken (ítem 18.9) y el segundo caso el límite inferior calculado en el IC fue de ,59 (ítem 18.1).

Tabla 10.
 Cálculos V de Aiken e IC - bloque 18

Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec
P	4,64	,91	,81	,96		P	4,50	,88	,76	,94		P	4,14	,79	,66	,87	
18.1A	4,50	,88	,76	,94	T	18.4A	4,64	,91	,81	,96	O	18.7A	3,79	,70	,57	,80	E
C	3,86	,71	,59*	,82		C	4,50	,88	,76	,94		C	3,93	,73	,60	,83	
P	4,57	,89	,79	,95		P	4,43	,86	,74	,93		P	3,50	,63	,49	,74	
18.2A	4,64	,91	,81	,96	O	18.5A	4,57	,89	,79	,95	O	18.8A	3,36	,59	,46	,71	E
C	4,50	,88	,76	,94		C	4,36	,84	,72	,91		C	3,57	,64	,51	,76	
P	4,64	,91	,81	,96		P	4,36	,84	,72	,91		P	4,64	,91	,81	,96	
18.3A	4,29	,82	,70	,90	O	18.6A	4,50	,88	,76	,94	O	18.9A	4,36	,84	,72	,91	T
C	4,43	,86	,74	,93		C	4,64	,91	,81	,96		C	3,50	,63	,49	,74	

P: Pertinencia, A: Adecuación, C: Claridad

IC LI: Int. de confianza límite inferior, IC LS: Int. de confianza límite superior

O: Mantener original, T: Indicador transformado, E: Indicador eliminado

* Valores fuera de los límites admisibles

4.6. Resultados de la dimensión técnica

La rúbrica incorpora una dimensión final para evaluar aspectos de índole técnica como la usabilidad, la accesibilidad o el nivel de compatibilidad. La Tabla 11 resume los resultados obtenidos en el bloque de indicadores 19, dedicado a medir el nivel de usabilidad de una app. Como se puede observar se tomó la decisión de eliminar 2 ítems (19.4 y 19.8), por presentar niveles de acuerdo entre evaluadores bajos en los tres aspectos evaluados (pertinencia, adecuación y claridad) tras la iteración final. Por otro lado, se modificó la redacción original de cuatro indicadores tras mantener niveles bajos de acuerdo en torno a la claridad (19.5, 19.6, 19.12 y 19.16).

Tabla 11.
 Cálculos V de Aiken e IC - bloques 19, 20 y 21

Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec	Ítem	M	VA	IC LI	IC LS	Dec
19.1	P 4,50 ,88 ,76 ,94	A 4,43 ,86 ,74 ,93	O 4,29 ,82 ,70 ,90			19.7	P 4,64 ,91 ,81 ,96	A 4,64 ,91 ,81 ,96	O 4,64 ,91 ,81 ,96			19.13	P 4,43 ,86 ,74 ,93	A 4,36 ,84 ,72 ,91	O 4,50 ,88 ,76 ,94		
19.2	P 4,50 ,88 ,76 ,94	A 4,43 ,86 ,74 ,93	O 4,43 ,86 ,74 ,93			19.8	P 3,50 ,63*,49*,74	A 3,71 ,68*,55*,79	E 3,64 ,66*,53*,77			19.14	P 4,71 ,93 ,83 ,97	A 4,29 ,82 ,70 ,90	O 4,29 ,82 ,70 ,90		
19.3	P 4,50 ,88 ,76 ,94	A 4,50 ,88 ,76 ,94	O 4,36 ,84 ,72 ,91			19.9	P 4,50 ,88 ,76 ,94	A 4,71 ,93 ,83 ,97	O 4,50 ,88 ,76 ,94			19.15	P 4,36 ,84 ,72 ,91	A 4,57 ,89 ,79 ,95	O 4,57 ,89 ,79 ,95		
19.4	P 3,57 ,64*,51*,76	A 3,64 ,66*,53*,77	E 3,79 ,70 ,57*,80			19.10	P 4,29 ,82 ,70 ,90	A 4,86 ,96 ,88 ,99	O 4,64 ,91 ,81 ,96			19.16	P 4,36 ,84 ,72 ,91	A 4,71 ,93 ,83 ,97	T 3,50 ,63*,49*,74		
19.5	P 4,29 ,82 ,70 ,90	A 4,43 ,86 ,74 ,93	T 4,00 ,75 ,62*,84			19.11	P 4,36 ,84 ,72 ,91	A 4,50 ,88 ,76 ,94	O 4,57 ,89 ,79 ,95			19.17	P 4,43 ,86 ,74 ,93	A 4,43 ,86 ,74 ,93	O 4,50 ,88 ,76 ,94		
19.6	P 4,57 ,89 ,79 ,95	A 4,71 ,93 ,83 ,97	T 4,21 ,80 ,68*,89			19.12	P 4,71 ,93 ,83 ,97	A 4,50 ,88 ,76 ,94	T 3,93 ,73 ,60*,83			19.18	P 4,43 ,86 ,74 ,93	A 4,43 ,86 ,74 ,93	O 4,50 ,88 ,76 ,94		
20.1	P 4,36 ,84 ,72 ,91	A 4,64 ,91 ,81 ,96	O 4,36 ,84 ,72 ,91			20.8	P 3,57 ,64*,51*,76	A 3,43 ,61*,48*,72	E 3,36 ,59*,46*,71			20.15	P 4,36 ,84 ,72 ,91	A 4,36 ,84 ,72 ,91	O 4,57 ,89 ,79 ,95		
20.2	P 4,43 ,86 ,74 ,93	A 4,50 ,88 ,76 ,94	O 4,57 ,89 ,79 ,95			20.9	P 4,29 ,82 ,70 ,90	A 4,57 ,89 ,79 ,95	O 4,43 ,86 ,74 ,93			20.16	P 4,64 ,91 ,81 ,96	A 4,43 ,86 ,74 ,93	O 4,50 ,88 ,76 ,94		
20.3	P 3,29 ,57*,44*,69	A 3,36 ,59*,46*,71	E 3,21 ,55*,42*,68			20.10	P 4,50 ,88 ,76 ,94	A 4,57 ,89 ,79 ,95	O 4,71 ,93 ,83 ,97			20.17	P 4,64 ,91 ,81 ,96	A 4,57 ,89 ,79 ,95	O 4,43 ,86 ,74 ,93		
20.4	P 4,57 ,89 ,79 ,95	A 4,57 ,89 ,79 ,95	O 4,43 ,86 ,74 ,93			20.11	P 4,71 ,93 ,83 ,97	A 4,64 ,91 ,81 ,96	O 4,71 ,93 ,83 ,97			20.18	P 3,93 ,73 ,60*,83	A 3,86 ,71 ,59*,82	E 3,64 ,66*,53*,77		
20.5	P 4,50 ,88 ,76 ,94	A 4,43 ,86 ,74 ,93	O 4,36 ,84 ,72 ,91			20.12	P 4,50 ,88 ,76 ,94	A 4,43 ,86 ,74 ,93	T 3,71 ,68*,55*,79			20.19	P 4,50 ,88 ,76 ,94	A 4,43 ,86 ,74 ,93	O 4,57 ,89 ,79 ,95		
20.6	P 4,71 ,93 ,83 ,97	A 4,43 ,86 ,74 ,93	O 4,50 ,88 ,76 ,94			20.13	P 4,43 ,86 ,74 ,93	A 4,50 ,88 ,76 ,94	O 4,50 ,88 ,76 ,94			20.20	P 4,43 ,86 ,74 ,93	A 4,29 ,82 ,70 ,90	T 3,79 ,70 ,57*,80		
20.7	P 4,50 ,88 ,76 ,94	A 4,36 ,84 ,72 ,91	T 4,07 ,77 ,64*,86			20.14	P 4,29 ,82 ,70 ,90	A 4,43 ,86 ,74 ,93	O 4,36 ,84 ,72 ,91								
21.1	P 4,29 ,82 ,70 ,90	A 4,57 ,89 ,79 ,95	O 4,50 ,88 ,76 ,94			21.4	P 4,64 ,91 ,81 ,96	A 4,21 ,80 ,68*,89	E 4,36 ,84 ,72 ,91			21.7	P 4,57 ,89 ,79 ,95	A 4,43 ,86 ,74 ,93	O 4,71 ,93 ,83 ,97		

P 4,29 ,82 ,70 ,90	P 4,43 ,86 ,74 ,93
21.2 A 4,50 ,88 ,76 ,94O	21.5 A 4,71 ,93 ,83 ,97T
C 4,50 ,88 ,76 ,94	C 4,21 ,80 ,68* ,89
P 4,43 ,86 ,74 ,93	P 4,64 ,91 ,81 ,96
21.3 A 4,50 ,88 ,76 ,94O	21.6 A 4,57 ,89 ,79 ,95O
C 4,71 ,93 ,83 ,97	C 4,43 ,86 ,74 ,93

P: Pertinencia, A: Adecuación, C: Claridad

IC LI: Int. de confianza límite inferior, IC LS: Int. de confianza límite superior

O: Mantener original, T: Indicador transformado, E: Indicador eliminado

* Valores fuera de los límites admisibles

Tras la última iteración del proceso, el bloque 20 que mide el nivel de accesibilidad de una app, quedó estructurado en un conjunto final de 17 indicadores de evaluación. Se puede observar que fue necesario eliminar 3 ítems que arrojaron valores bajos del estadístico V tanto a nivel de pertinencia como de adecuación (20.3 y 20.8), así como del LI en el ítem 20.18. Además, se reformularon 3 indicadores en los que no se superó el umbral establecido de ,7 en la dimensión de claridad (20.7, 20.12 y 20.20) Por último, el bloque 21 se destinó a evaluar el nivel de compatibilidad que ofrecen las Apps. Los resultados obtenidos condujeron a eliminar uno de los indicadores propuestos (21.4) por considerarse poco adecuado para la dimensión (LI=,68). También se modificó la redacción del ítem 21.5 (LI=,68 en la dimensión de claridad).

5. Discusión y conclusiones

La digitalización de la sociedad actual es un fenómeno tan evidente como imparable, por lo que cada vez más ámbitos de conocimientos y acciones se realizan en entornos digitales. La educación formal no es ajena a este fenómeno empleando cada vez más herramientas, entornos virtuales y tecnologías móviles para proveer de objetos de conocimiento y actividades formativas al alumnado (Brazuelo-Grund et al., 2017; Kearney et al., 2019). Si bien estos procesos están orientados y usualmente regulados por profesionales educativos, éstos no siempre poseen las capacidades y formación necesaria para enjuiciar adecuadamente el valor educativo de muchas de las herramientas de las que se dispone (Almerich et al., 2018). Tanto en ámbitos formales como informales, los jóvenes emplean herramientas digitales y App para apoyar su aprendizaje, relacionarse con los demás u ocupar sus espacios de ocio (Symeonaki et al., 2022), pero suelen ser tiempos y espacios autorregulados por el propio joven, que ni posee la competencia ni la capacidad de valorar educativamente las herramientas que emplea.

Investigaciones recientes destacan la facilidad de acceso a la información y un mayor compromiso con el aprendizaje como los principales impactos de la tecnología móvil en el aula, indicando también que la elección de las aplicaciones móviles a utilizar depende en gran medida de la percepción de los docentes sobre el impacto de estas aplicaciones en el aprendizaje (Gómez-Domingo & Badia-Garganté, 2016; Wai et al., 2018), una percepción en ocasiones subjetiva o que no responde a un proceso de reflexión riguroso y estructurado.

En la literatura científica podemos encontrar diferentes aproximaciones y marcos pedagógicos que han sido propuestos para el aprendizaje móvil (Hsu & Ching, 2015) que enmarcan los procesos de evaluación desde posicionamientos distintos. Algunos estudios prestan mayor atención a la evaluación de componentes técnicos, aun explorando también el nivel de eficiencia y funcionalidad de una App (Cherner et al., 2016). Otras propuestas como la de Papadakis et al. (2017), consideran aspectos evaluativos como la provisión de aprendizajes, la motivación, el feedback y la monitorización del progreso. Algunas aproximaciones analizan las características pedagógicas de las App, como la colaboración y compartición de información, las posibilidades de personalización y la autenticidad (Burden & Kearney, 2018; Kearney et al., 2019), o se centran en la experiencia de aprendizaje del usuario (Lee & Cherner, 2015), pero no consideran los procesos psicológicos subyacentes. La mayor parte de los estudios previos encontrados enfatizan las dimensiones de evaluación vinculadas al diseño, la interacción App-usuario o la seguridad frente a aquellas que se centran en cuestiones educativas (Shoukry et al., 2015; Sweeney & Moore, 2012). Otros estudios se focalizan en contextos educativos específicos, como es el caso de la educación especial (Walker, 2011; Weng, 2015).

Estos modelos de evaluación, abordan aspectos importantes a tener en cuenta en el proceso de toma de decisiones en torno a la selección de una App para un contexto de enseñanza-aprendizaje, pero en dichos modelos se presta mayor atención a aspectos técnicos y no se consideran de manera específica algunos procesos clave vinculados al aprendizaje desde los ámbitos personal, cognitivo-metacognitivo, o social (Abenti, 2020; Berrios-Aguayo et al., 2021; Expósito-López, 2018; Expósito-López & Olmedo-Moreno, 2020; Harrison, 2009).

Este estudio ha permitido dar forma un nuevo instrumento, en forma de rúbrica, para enjuiciar el valor educativo de las App, al que se ha denominado E-V-A (Educational Value Assesment). Para ello, una primera fase de diseño conceptual y teórico de la rúbrica dio lugar a un conjunto de dimensiones y sub-dimensiones desde el ámbito teórico, tras lo cual se construyó el inventario inicial de ítems que posteriormente fue sometido a revisión empleando un proceso de Delphi con tres vueltas en el que participaron 14 jueces evaluadores expertos.

El proceso de validación de contenido apoyado en el cálculo del Coeficiente de Correlación Intraclase y del Coeficiente de validez de contenido V de Aiken, permitió estructurar la rúbrica en un conjunto final de cinco dimensiones y un total de 100 ítems de evaluación, que abarcan los elementos clave del desarrollo de los procesos educativos.

El instrumento generado considera un conjunto de variables educativas intrínsecas que permiten analizar si una App puede mejorar o ayudar a desarrollar la personalidad del usuario, si una App puede ser una herramienta adecuada para conseguir la mejora del proceso de aprendizaje del usuario y si puede contribuir a la interacción social del usuario. Además, también considera otras variables educativas ligadas al soporte tecnológico como son la accesibilidad, la interacción usuario-máquina y la provisión de feedback educativo. En el anexo se puede consultar la versión resumida de la rúbrica.

La rúbrica puede facilitar la labor a educadores y docentes proporcionando un conjunto completo de criterios, fundamentado en distintos ámbitos epistemológicos, para evaluar el valor educativo de una aplicación para dispositivos móviles. El protocolo de evaluación servirá para que los profesionales del ámbito educativo, tanto en la educación superior como también en etapas educativas pre-universitarias, dispongan de un procedimiento ordenado y estructurado en base al cual tomar decisiones de forma crítica a la hora de utilizar una determinada App en un contexto educativo o situación de enseñanza - aprendizaje.

Teniendo en cuenta que, como estudio preliminar de validación, la rúbrica generada no se ha sometido a un proceso de ensayo generalizado, la primera actuación prospectiva que se deriva de este trabajo es la realización de un proyecto piloto de testeo de la rúbrica, sometiendo a evaluación un conjunto de Apps en base a diferentes contextos de intervención educativa que se planteen. Por otro lado, planteamos la posibilidad de someter esta versión de la rúbrica a una nueva evaluación de descriptores de evaluación, utilizando también para ello un proceso de juicio de expertos, que puedan incluso considerar aspectos relevantes de las pruebas piloto.

Presentación del artículo: 10 de octubre de 2022

Fecha de aprobación: 13 de diciembre 2022

Fecha de publicación: 9 de enero de 2023

Expósito-López, J., Romero-Díaz de la Guardia, J. J., & Moreno-Olmedo, E. M. (2023). Diseño y validación de contenido de una rúbrica para medir el valor educativo de aplicaciones para dispositivos móviles. *RED. Revista de Educación a Distancia*, 23(72). <http://dx.doi.org/10.6018/red.542261>

Financiación

Trabajo financiado como parte de un Proyecto I+D+I para la realización de «pruebas de concepto», en el marco del programa estatal de I+D+I orientada a los retos de la sociedad, del plan estatal de investigación científica y técnica y de innovación 2017-2020. Convocatoria: 2021. Referencia: PDC2021-121306-I00. Título: Herramientas eficaces para la integración escolar y social: Educational Value Assessment y App STUDENG+ (STUDENG+).

Referencias

Abenti, H. F. (2020). How do I teach you? An examination of multiple intelligences and the impact on communication in the classroom. *Language & Communication*, 73, 29-33. <https://doi.org/10.1016/j.langcom.2020.04.001>

Almerich, G., Díaz-García, I., Cebrián-Cifuentes, S., & Suárez-Rodríguez, J. (2018). Estructura dimensional de las competencias del siglo XXI en los estudiantes universitarios de educación. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 24(1), Art. 1. <https://doi.org/10.7203/relieve.24.1.12548>

- Arteaga-Martínez, B., Macías-Sánchez, J., Pla-Castells, M., & Ramírez-García, M. (2021). Diseño y validación de un instrumento para observación de clases de matemáticas en Educación Secundaria: Grupo nominal y método Delphi. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, 27(2), Art. 2. <https://doi.org/10.30827/relieve.v27i2.21812>
- Berrios-Aguayo, ., Arazola-Ruano, C., & Pantoja-Vallejo, A. (2021). Multiple intelligences: Educational and cognitive development with a guiding focus. *South African Journal of Education*, 41(2). Scopus. <https://doi.org/10.15700/saje.v41n2a1828>
- Brazuelo-Grund, F., Gallego-Gil, D. J., & Cacheiro-González, M. L. (2017). Los docentes ante la integración educativa del teléfono móvil en el aula. *Revista de Educación a Distancia (RED)*, 52, Art. 52. <https://revistas.um.es/red/article/view/282201>
- Burden, K., & Kearney, M. (2018). Designing an Educator Toolkit for the Mobile Learning Age. *International Journal of Mobile and Blended Learning (IJMBL)*, 10(2), 88-99. <https://doi.org/10.4018/IJMBL.2018040108>
- Cabero-Almenara, J., & Barroso-Osuna, J. (2013). La utilización del juicio de experto para la evaluación de TIC: El Coeficiente de Competencia Experta. *Bordón. Revista de Pedagogía*, 65(2), Art. 2.
- Cabero-Almenara, J., & Llorente-Cejudo, M. C. (2013). La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC). *Eduweb*, 7(2), 11-22.
- Camilleri, M. A., & Camilleri, A. C. (2019). The students' readiness to engage with mobile learning apps. *Interactive Technology and Smart Education*, 17(1), 28-38. <https://doi.org/10.1108/ITSE-06-2019-0027>
- Charter, R. A. (2003). A Breakdown of Reliability Coefficients by Test Type and Reliability Method, and the Clinical Implications of Low Reliability. *The Journal of General Psychology*, 130(3), 290-304. <https://doi.org/10.1080/00221300309601160>
- Chau, C. L. (2014). Positive Technological Development for Young Children in the Context of Children's Mobile Apps [Ph.D., Tufts University]. En *ProQuest Dissertations and Theses*. <https://www.proquest.com/docview/1551196695/abstract/91A9D8FA5E03478CPQ/1>
- Cherner, T. S., Lee, C. H., Fegely, A., & Santaniello, L. A. (2016). A Detailed Rubric for Assessing the Quality of Teacher Resource Apps. *Journal of Information Technology Education: Innovations in Practice*, 15, 117-143.
- Duque, M. P., & Packer, M. J. (2014). Pensamiento y lenguaje. El proyecto de Vygotsky para resolver la crisis de la Psicología. *Tesis Psicológica*, 9(2), 30-57.

- European Commission. (2021). *Digital Education Action Plan*. https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_es
- European Parliament. (2016). Directiva (UE) 2016/2102 del Parlamento Europeo y del Consejo, de 26 de octubre de 2016, sobre la accesibilidad de los sitios web y aplicaciones para dispositivos móviles de los organismos del sector público. *Official Journal of the European Union*, 59, 1-15.
- Expósito-López, J. (2018). *Acción tutorial en la universidad*. Comares.
- Expósito-López, J., & Olmedo-Moreno, E. M. (2020). Scientometric analysis of publications on guidance, mentoring, and tutorial action registered on databases. *Formación universitaria*, 13(3), 123-138. <https://doi.org/10.4067/S0718-50062020000300123>
- Gómez-Domingo, M., & Badia-Garganté, A. (2016). Exploring the use of educational technology in primary education: Teachers' perception of mobile technology learning impacts and applications' use in the classroom. *Computers in Human Behavior*, 56, 21-28. <https://doi.org/10.1016/j.chb.2015.11.023>
- Harrison, R. (2009). *Learning and development*. Chartered Institute of Personnel and Development.
- Hsu, Y. C., & Ching, Y. H. (2015). A review of models and frameworks for designing mobile learning experiences and environments. *Canadian Journal of Learning and Technology*, 41(3), Art. 3. <https://doi.org/10.21432/T2V616>
- Kalimaya, S., Feranie, S., & Agustin, R. R. (2021). *The effect of multiple intelligence theory based teaching towards students' achievement on electrical circuit topic*. 1806(1). Scopus. <https://doi.org/10.1088/1742-6596/1806/1/012131>
- Kearney, M., Burden, K., & Schuck, S. (2019). Disrupting Education Using Smart Mobile Pedagogies. En L. Daniela (Ed.), *Didactics of Smart Pedagogy: Smart Pedagogy for Technology Enhanced Learning* (pp. 139-157). Springer International Publishing. https://doi.org/10.1007/978-3-030-01551-0_7
- 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>
- Krathwohl, D. R., & Anderson, L. W. (2010). Merlin C. Wittrock and the Revision of Bloom's Taxonomy. *Educational Psychologist*, 45(1), 64-65. <https://doi.org/10.1080/00461520903433562>
- Lee, C. Y., & Cherner, T. S. (2015). A Comprehensive Evaluation Rubric for Assessing Instructional Apps. *Journal of Information Technology Education: Research*, 14, 21-53.

- Li, K. C., Lee, L. Y.-K., Wong, S.-L., Yau, I. S.-Y., & Wong, B. T.-M. (2018). Effects of mobile apps for nursing students: Learning motivation, social interaction and study performance. *Open Learning*, 33(2), 99-114. <https://doi.org/10.1080/02680513.2018.1454832>
- Merino-Soto, C., & Livia-Segovia, J. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales de Psicología / Annals of Psychology*, 25(1), Art. 1.
- Real Decreto 1112/2018, de 7 de septiembre, sobre accesibilidad de los sitios web y aplicaciones para dispositivos móviles del sector público, Pub. L. No. Real Decreto 1112/2018, BOE-A-2018-12699 90533 (2018). <https://www.boe.es/eli/es/rd/2018/09/07/1112>
- Morales, J. C., Ramírez, N. E., Vargas, S. H., Peñuela, A. J., Morales, J. C., Ramírez, N. E., Vargas, S. H., & Peñuela, A. J. (2020). Use of Mobile apps in the classroom and its determining factors. *Formación universitaria*, 13(6), 13-22. <https://doi.org/10.4067/S0718-50062020000600013>
- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, 60(5), 969-983. <https://doi.org/10.1002/asi.21030>
- Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2017). Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies*, 22(6), 3147-3165. <https://doi.org/10.1007/s10639-017-9579-0>
- Penfield, R. D., & Giacobbi, P. R. (2004). Applying a Score Confidence Interval to Aiken's Item Content-Relevance Index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. https://doi.org/10.1207/s15327841mpee0804_3
- Puentedura, R. R. (2014). SAMR and TPCK: A Hands-On Approach to Classroom Practice. *Ruben R. Puentedura's Weblog. Ongoing thoughts on education and technology*. <http://www.hippasus.com/rrpweblog/archives/000140.html>
- Salazar-Gómez, E., Tobón, S., & Juárez-Hernández, L. G. (2018). Diseño y validación de una rúbrica de evaluación de las competencias digitales desde la socioformación. *Apuntes Universitarios*, 8(3), Art. 3. <https://doi.org/10.17162/au.v8i3.329>
- Shoukry, L., Sturm, C., & Galal-Edeen, G. (2015). *Pre-MEGa: A Proposed Framework for the Design and Evaluation of Preschoolers' Mobile Educational Games*. The International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 12), Bridgeport: USA. https://doi.org/10.1007/978-3-319-06773-5_52
- Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015). Mobile App Rating Scale: A New Tool for Assessing the Quality of

Health Mobile Apps. *JMIR MHealth and UHealth*, 3(1), e3422.
<https://doi.org/10.2196/mhealth.3422>

Sweeney, P., & Moore, C. (2012). Mobile Apps for Learning Vocabulary: Categories, Evaluation and Design Criteria for Teachers and Developers. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 2(4), 1-16.
<https://doi.org/10.4018/ijcallt.2012100101>

Symeonaki, M., Filandrianos, G., & Stamou, G. (2022). Visualising key information and communication technologies (ICT) indicators for children and young individuals in Europe. *Humanities and Social Sciences Communications*, 9(1). Scopus.
<https://doi.org/10.1057/s41599-022-01356-5>

Tabuenca, B., Sánchez-Peña, J. J., & Cuetos-Revuelta, M. J. (2019). El smartphone desde la perspectiva docente: ¿una herramienta de tutorización o un catalizador de ciberacoso? *Revista de Educación a Distancia (RED)*, 19(59), Art. 59.
<https://doi.org/10.6018/red/59/01>

Taylor, G., Kolak, J., Bent, E. M., & Monaghan, P. (2022). Selecting educational apps for preschool children: How useful are website app rating systems? *British Journal of Educational Technology*, 53(5), 1262-1282. <https://doi.org/10.1111/bjet.13199>

UNESCO. (2019). *Tesaurus de la UNESCO (Educación)*.
<https://vocabularies.unesco.org/browser/thesaurus/es/page/domain1>

Wai, I. S. H., Ng, S. S. Y., Chiu, D. K. W., Ho, K. K. W., & Lo, P. (2018). Exploring undergraduate students' usage pattern of mobile apps for education. *Journal of Librarianship and Information Science*, 50(1), 34-47.
<https://doi.org/10.1177/0961000616662699>

Walker, H. (2011). Evaluating the effectiveness of apps for mobile devices. *Journal of Special Education Technology*, 26(4), 59-63.
<https://doi.org/10.1177/016264341102600405>

Weng, P. L. (2015). Developing an App Evaluation Rubric for Practitioners in Special Education. *Journal of Special Education Technology*, 30(1), 43-58.
<https://doi.org/10.1177/016264341503000104>

Zapata-Ros, M. (2012). Calidad en entornos ubicuos de aprendizaje. *Revista de Educación a Distancia (RED)*, 31, Art. 31.
<https://revistas.um.es/red/article/view/232871>

Anexo. Rúbrica Educational Value Assessment E-V-A (versión resumida)

DIMENSIÓN DE DESARROLLO PERSONAL						
El uso de la app ayuda al usuario a desarrollar aspectos referidos a la personalidad del usuario						
-3	-2	-1	0	+1	+2	+3
La app de forma constante envía al usuario mensajes negativos sobre su apariencia o personalidad	La app de forma frecuente envía al usuario mensajes negativos sobre su apariencia o personalidad	La app en ocasiones envía al usuario mensajes negativos sobre su apariencia o personalidad	No aplicable	La app en ocasiones envía al usuario mensajes positivos sobre su apariencia o personalidad	La app de forma frecuente envía al usuario mensajes positivos sobre su apariencia o personalidad	La app de forma constante envía al usuario mensajes negativos sobre su apariencia o personalidad
DIMENSIÓN DESARROLLO PERSONAL						
El uso de la app facilita al usuario el desarrollo de la inteligencia emocional: Competencias para el bienestar						
-3	-2	-1	0	+1	+2	+3
La app dificulta de forma constante que el usuario pueda fijar sus propias metas, expresar su opinión y desarrollar hábitos adecuados para su vida	La app dificulta de forma frecuente que el usuario pueda fijar sus propias metas, expresar su opinión y desarrollar hábitos adecuados para su vida	La app dificulta en algunas ocasiones que el usuario pueda fijar sus propias metas, expresar su opinión y desarrollar hábitos adecuados para su vida	No aplicable	La app promueve en algunas ocasiones que el usuario pueda fijar sus propias metas, expresar su opinión y desarrollar hábitos adecuados para su vida	La app promueve de forma frecuente que el usuario pueda fijar sus propias metas, expresar su opinión y desarrollar hábitos adecuados para su vida	La app promueve de forma constante que el usuario pueda fijar sus propias metas, expresar su opinión y desarrollar hábitos adecuados para su vida
DIMENSIÓN DE DESARROLLO COGNITIVO Y METACOGNITIVO						
El uso de la app requiere que el usuario emplee y desarrolle estrategias cognitivas y metacognitivas						
-3	-2	-1	0	+1	+2	+3
La app dificulta de forma constante o no promueve el desarrollo de estrategias cognitivas como el ensayo de opinión, la organización o la elaboración de informaciones. Y meta-cognitivas para organizar el propio pensamiento	La app dificulta de forma frecuente o no promueve el desarrollo de estrategias cognitivas como el ensayo de opinión, la organización o la elaboración de informaciones. Y meta-cognitivas para organizar el propio pensamiento	La app dificulta en ocasiones o no promueve el desarrollo de estrategias cognitivas como el ensayo de opinión, la organización o la elaboración de informaciones. Y meta-cognitivas para organizar el propio pensamiento	No aplicable	La app fomenta en ocasiones o propicia el desarrollo de estrategias cognitivas como el ensayo de opinión, la organización o la elaboración de informaciones. Y meta-cognitivas para organizar el propio pensamiento	La app fomenta de forma frecuente o propicia el desarrollo de estrategias cognitivas como el ensayo de opinión, la organización o la elaboración de informaciones. Y meta-cognitivas para organizar el propio pensamiento	La app fomenta de forma constante o propicia el desarrollo de estrategias cognitivas como el ensayo de opinión, la organización o la elaboración de informaciones. Y meta-cognitivas para organizar el propio pensamiento
DIMENSIÓN DE DESARROLLO DEL APRENDIZAJE Y DE COMPETENCIAS						
El uso de la app fomenta el proceso general de aprendizaje del usuario y el desarrollo de competencias para aprender						
-3	-2	-1	0	+1	+2	+3
La app desmotiva al usuario a aprender nuevos contenidos con valor educativo, y fomenta negativamente el uso de competencias para aprender (lógico matemáticas, de comunicación artísticas o físicas) adecuadamente	La app de forma frecuente de forma constante desmotiva al usuario a aprender nuevos contenidos con valor educativo, y fomenta negativamente el uso de la mayoría de las competencias para aprender (lógico matemáticas, de comunicación artísticas o físicas) adecuadamente	La app en ocasiones desmotiva al usuario a aprender nuevos contenidos con valor educativo, y fomenta negativamente algunas competencias para aprender (lógico matemáticas, de comunicación artísticas o físicas) adecuadamente	No aplicable	La app en ocasiones motiva al usuario a aprender nuevos contenidos con valor educativo, y fomenta positivamente algunas competencias para aprender (lógico matemáticas, de comunicación artísticas o físicas) adecuadamente	La app de forma frecuente motiva al usuario a aprender nuevos contenidos con valor educativo, y fomenta positivamente la mayoría de las competencias para aprender (lógico matemáticas, de comunicación artísticas o físicas) adecuadamente	La app de forma constante motiva muy positivamente al usuario a aprender nuevos contenidos con valor educativo, y fomenta positivamente todas las competencias para aprender (lógico matemáticas, de comunicación artísticas o físicas) adecuadamente

DIMENSIÓN DEL DESARROLLO SOCIAL						
El uso de la app fomenta el desarrollo de aspectos referidos a las habilidades sociales e integración social de los usuarios						
-3	-2	-1	0	+1	+2	+3
La app de forma constante fomenta un conocimiento negativo y el desarrollo de comportamientos en contra del mundo físico-natural, social y cultural	La app de forma frecuente fomenta un conocimiento poco adecuado y el desarrollo de algunos comportamientos en contra del mundo físico-natural, social y cultural	La app en ocasiones fomenta un conocimiento poco adecuado y el desarrollo de algunos comportamientos en contra del mundo físico-natural, social y cultural	No aplicable	La app en ocasiones fomenta un conocimiento adecuado y el desarrollo de algunos comportamientos a favor del mundo físico-natural, social y cultural	La app de forma frecuente fomenta un conocimiento adecuado y el desarrollo de algunos comportamientos a favor del mundo físico-natural, social y cultural	La app de forma constante fomenta un conocimiento positivo y el desarrollo de comportamientos a favor del mundo físico-natural, social y cultural
DIMENSIÓN TÉCNICA – USABILIDAD						
La app puede ser utilizada por los usuarios de forma satisfactoria						
-3	-2	-1	0	+1	+2	+3
La app presenta importantes deficiencias que dificultan la usabilidad, siendo necesario un replanteamiento general de los siguientes aspectos: diseño, despliegue de contenidos, navegación o gestión de la retroalimentación al usuario	La app presenta un nivel de usabilidad bajo, requiriendo mejorar en todos los siguientes aspectos: diseño, despliegue de contenidos, navegación o gestión de la retroalimentación al usuario	La app tiene un nivel de usabilidad inferior a lo esperado, siendo necesario mejorar al menos tres de los siguientes aspectos: diseño, despliegue de contenidos, navegación o gestión de la retroalimentación al usuario	No aplicable	La app tiene un nivel de usabilidad general adecuado, aunque necesita mejorar en más de uno de estos aspectos: diseño, despliegue de contenidos, navegación o gestión de la retroalimentación al usuario	La app tiene un buen nivel de usabilidad, aunque puede ser mejorable en alguno de estos aspectos: diseño, despliegue de contenidos, navegación o gestión de la retroalimentación al usuario	La app tiene un nivel de usabilidad óptimo en términos de diseño, despliegue de contenidos, navegación y gestión de la retroalimentación al usuario
DIMENSIÓN TÉCNICA – ACCESIBILIDAD						
La app puede ser utilizada sin que haya barreras que dificulten la interacción o el acceso a sus elementos						
-3	-2	-1	0	+1	+2	+3
El diseño de la app limita en gran medida la accesibilidad, siendo necesario ofrecer alternativas en todos los siguientes aspectos: Tipografía y colores, formato de presentación de la información, gestión de la retroalimentación e interacción del usuario con la app	El diseño de la app limita bastante la accesibilidad en al menos tres de los siguientes aspectos: Tipografía y colores, formato de presentación de la información, gestión de la retroalimentación e interacción del usuario con la app	El diseño de la app limita la accesibilidad en al menos uno de los siguientes aspectos: Tipografía y colores, formato de presentación de la información, gestión de la retroalimentación e interacción del usuario con la app	No aplicable	El diseño de la app presenta un buen nivel de accesibilidad, aunque podría mejorar ofreciendo más alternativas para al menos 2 de los siguientes aspectos: Tipografía y colores, formato de presentación de la información, retroalimentación o interacción con la app	El diseño de la app presenta un buen nivel de accesibilidad, aunque podría mejorar ofreciendo más alternativas para alguno de estos aspectos: Tipografía y colores, formato de presentación de la información, retroalimentación o interacción con la app	El diseño de la app es óptimo en términos de accesibilidad, ofreciéndose suficientes alternativas para la gestión tipográfica y de colores, el formato de presentación de la información, la retroalimentación al usuario y la interacción con la app