



## Validez de contenido de indicadores para evaluar la calidad de aplicaciones móviles dirigidas a menores que promueven la educación en salud a través de la gamificación

### *Content Validity of Indicators for Evaluating the Quality of Mobile Applications Aimed at Children that Promote Health Education through Gamification*

  Claudio Delgado-Morales (C.D.M.). Universidad de Huelva (España)

  Ana Duarte-Hueros (A.D.H.). Universidad de Huelva (España)

#### RESUMEN

El uso extendido de dispositivos móviles en menores ha impulsado el desarrollo de aplicaciones móviles destinadas a promover la educación en salud, muchas de ellas con elementos de *gamificación* integrados como parte de sus funcionalidades. No obstante, la falta de marcos de evaluación comunes y criterios de calidad afectan a la fiabilidad de estos recursos digitales. Por ello, el objetivo principal de este estudio es validar a nivel de contenido, mediante la técnica del juicio de expertos y expertas, un conjunto de 92 indicadores que sirvan de base para construir un instrumento de evaluación integral. Han sido seleccionadas un total de 20 personas para formar parte del panel, el cual se ha caracterizado por la diversidad disciplinar y la trayectoria profesional. Los resultados han revelado que los índices y estadísticos analizados, en líneas generales, han sido buenos-excelentes con una alta concordancia interjueces. Aunque algunos indicadores han sido modificados, reubicados o eliminados por no alcanzar valores aceptables, finalmente se ha establecido un listado de 72 indicadores susceptibles de ser materializados en un instrumento global que sirva de orientación en el diseño, el desarrollo y la evaluación de este tipo de aplicaciones móviles educativas que incorporan contenidos relacionados con la salud.

#### ABSTRACT

*The widespread use of mobile devices among minors has driven the development of mobile applications aimed at promoting health education, many of which incorporate gamification elements as part of their functionalities. However, the lack of common evaluation frameworks and quality criteria undermines the reliability of these digital resources. Therefore, the main objective of this study is to validate, through expert judgment, a set of 92 content indicators as the basis for building a comprehensive evaluation instrument. A total of 20 professionals were selected to form a panel characterized by disciplinary diversity and professional experience. The results revealed that, overall, the analyzed indices and statistics showed good to excellent values, with high inter-rater agreement. Although some indicators were modified, relocated, or eliminated due to unacceptable scores, a final set of 72 indicators was established. These may be used to develop a comprehensive tool to guide the design, development, and evaluation of educational mobile applications that include health-related content.*

#### PALABRAS CLAVE - KEYWORDS

Validez de contenido, Calidad, Aplicaciones móviles, Hábitos saludables, Menores  
*Content validity, Quality, Mobile applications, Healthy habits, Children*



## 1. INTRODUCCIÓN

La integración de la tecnología móvil en prácticamente la totalidad de ámbitos y sectores poblacionales cada vez es mayor, pues ha ido progresivamente transformando la manera de entretenerse, trabajar, aprender, organizar y llevar a cabo tareas de la vida cotidiana, etc.

En concreto, según el Instituto Nacional de Estadística (INE, 2024), el 69,6% de menores en España entre 10 y 15 años disponían de un teléfono móvil en el 2024. En este sentido, niños, niñas y adolescentes, en contextos formales e informales, pasan numerosas horas de su tiempo interactuando con las aplicaciones móviles (*apps*) al impulsar factores asociados a la motivación y la diversión (Brown et al., 2022; Expósito-López et al., 2023).

En los últimos años, han emergido exponencialmente *apps* que incorporan contenidos relacionados con la salud. Este tipo de *apps* forman parte de la denominada salud móvil (del término en inglés *mHealth*), referida a la gestión de la atención médica y la salud pública con apoyo de dispositivos móviles como smartphones o tablets entre otros (World Health Organization [WHO], 2011). Precisamente, los dispositivos móviles son ya ampliamente utilizados por la población más joven en la promoción de la salud (Merino-Godoy et al., 2022).

Como exponen Alexandrou et al. (2023) y Thornton et al. (2021), las *apps* del ámbito de la *mHealth* ofrecen grandes posibilidades para involucrar a menores en procesos de cambios de comportamiento con respecto a su salud, especialmente en cuestiones sobre adquisición de hábitos saludables. Además, muchas de ellas integran técnicas de *gamificación* como insignias, puntos, niveles o tablas de clasificación para aumentar el compromiso y la participación (Fijačko et al., 2020; Jeminiwa et al., 2019; Sousa et al., 2020).

No obstante, la cantidad de *apps* educativas pertenecientes a la *mHealth* disponibles actualmente en el mercado, incluidas aquellas cuyas capacidades operativas y características técnicas han sido diseñadas y desarrolladas a partir de inteligencia artificial (IA) (Sefa-Yeboah et al., 2021), no garantiza una calidad proporcional en términos de fiabilidad y coherencia de la información contenida basada en evidencia científica, lo que dificulta tanto la identificación de un instrumento de evaluación pertinente e integral como la posterior selección de una *app* adecuada (Azad-Khaneghah et al., 2021; Cheng et al., 2020; Palacios-Gálvez et al., 2020).

Dado que los procesos y protocolos de regulación de calidad de estos recursos digitales son limitados a día de hoy y no existe un marco común de evaluación con un enfoque dinámico y multidimensional que pondere como corresponde las particularidades distintivas de cada *app* (Expósito-López et al., 2023; Tarricone et al., 2022), la población infantil y adolescente podría interiorizar patrones negativos de conducta para la salud al utilizar *apps* cuyos contenidos no están avalados ni validados por un conocimiento experto (Frontini et al., 2020; Ho et al., 2022).

En vista de la amplia disponibilidad de este tipo de software y la brecha actual en la revisión de estándares de calidad tan específicos, este estudio tiene como objetivo principal determinar la validez de contenido, mediante la técnica del juicio de expertos y expertas, de indicadores susceptibles de ser materializados en un instrumento para evaluar la calidad de las *apps* gamificadas que promueven la adquisición de hábitos saludables en menores. De esta forma, los sujetos participantes han valorado la adecuación de cada indicador a la subdimensión y dimensión de procedencia teniendo en cuenta la relevancia, la representatividad y la claridad como categorías de análisis.

Según autores y autoras como Ayre y Scally (2014), Escobar-Pérez y Cuervo-Martínez (2008), y Rubio et al. (2003), el juicio de expertos y expertas, como práctica generalizada en la comunidad científica y en investigaciones educativas, se trata de una opinión sistemática e informada de un conjunto de personas con experiencia reconocida

en el tema objeto de estudio, con capacidad suficiente para proporcionar información sólida basada en evidencia, así como valoraciones confiables, sobre la importancia de los indicadores y/o ítems dentro de un determinado instrumento de medición.

## 2. MÉTODO

### 2.1. Diseño del estudio

Siguiendo una metodología de juicio de expertos y expertas, se ha llevado a cabo un estudio descriptivo con diseño instrumental con el propósito de validar teóricamente, a nivel de contenido, una relación de indicadores que podrán concretarse y operativizarse en un instrumento de evaluación de la calidad de las *apps* en cuestión.

### 2.2. Indicadores evaluables y estructura preliminar del instrumento

Los indicadores de este estudio de validación han sido extraídos de la revisión sistemática de la literatura propuesta por Delgado-Morales y Duarte-Hueros (2023), quienes identificaron un listado actualizado e integral de dimensiones, subdimensiones e indicadores comunes en instrumentos de evaluación de la calidad de *apps* orientadas a la promoción de hábitos saludables en la ciudadanía. Además, se ha realizado una nueva consulta bibliográfica a fin de incluir indicadores y/o ítems recientes. Puede consultar el listado completo de dimensiones, subdimensiones e indicadores en Delgado-Morales y Duarte-Hueros (2025).

Dicho listado, que podrá derivar en el futuro instrumento de evaluación de este tipo de software, se compone de 92 indicadores asociados a 9 dimensiones principales: (1) Diseño gráfico y visual de la interfaz, con 6 indicadores; (2) Aspectos tecnológicos, con 7 indicadores; (3) Usabilidad de la *app*, con 8 indicadores; (4) Accesibilidad de la *app*, con 4 indicadores; (5) Información, con 21 indicadores; (6) Protección de datos, con 21 indicadores; (7) Gamificación, con 14 indicadores; (8) Efectos del uso y resultados, con 6 indicadores; y (9) Satisfacción, con 5 indicadores.

### 2.3. Selección de expertos y expertas

La selección del panel de expertos y expertas se ha realizado mediante un muestreo no probabilístico e intencional, donde los sujetos participantes han aceptado voluntariamente participar y conformar el panel. Al respecto, se ha recurrido a la accesibilidad en lugar de a la aleatoriedad, priorizando las características de los expertos y las expertas para el interés y el propósito del estudio de validación (McMillan y Schumacher, 2005).

En cuanto al número de participantes, siguiendo las recomendaciones de Hyrkäs et al. (2003), Rubio et al. (2003) y Witkin y Altschuld (1995), quienes sugieren un rango de entre 6 y 35 participantes para garantizar una estimación confiable de la validez de contenido.

Para la inclusión, se han aplicado los criterios propuestos por Davis (1992) y Skjong y Wentworth (2001), que contemplan: capacidad suficiente para tomar decisiones y emitir juicios rigurosos, prestigio a nivel científico, objetividad, y disponibilidad y compromiso de participación.

Se optó por el método de agregación individual (Violini et al., 2022), obteniendo información independiente de cada experto o experta, sin requerir contacto entre los sujetos participantes. No obstante, se han definido 5 grupos teniendo en cuenta el área de conocimiento o el ámbito profesional: (1) especialistas en el desarrollo de *apps* y/o software educativo e informática en general; (2) docentes de diferentes niveles educativos (educación primaria, educación secundaria obligatoria y educación superior); (3) profesionales del ámbito de la salud (medicina, enfermería, nutrición, psicología, etc.); (4) especialistas en métodos de investigación; y (5) personas expertas en tecnología educativa.

Una vez conformado, el panel de expertos y expertas se ha caracterizado por una gran diversidad disciplinar y trayectoria profesional. En total, han participado en la validación de contenido 20 personas: 12 mujeres y 8 hombres con un rango de edad 28-45 años ( $\bar{x}$ =38.4 años). De los 20 sujetos, 16 cuentan con estudios universitarios en las áreas de conocimiento de Ciencias de la Educación, Ciencias de la Salud e Ingeniería y, en concreto, 5 de ellos poseen el Grado de Doctor o Doctora, los cuales desempeñan su actividad docente e investigadora en universidades de España y Portugal. Por su parte, los sujetos con formación no universitaria disponen de la titulación de Ciclo Formativo de Grado Superior en el ámbito de la informática en general, ingeniería del software, desarrollo de aplicaciones web y multiplataforma, así como ciberseguridad. Sobre los años de experiencia, los expertos y las expertas presentan una media de 13.3 años.

## 2.4. Proceso de validación

Se ha contactado vía correo electrónico con cada sujeto participante, dando todos ellos su aceptación para formar parte del panel de expertos y expertas. Posteriormente, se les ha remitido la convocatoria formal.

Las valoraciones se han recabado por medio de una plantilla de evaluación, con una escala tipo Likert de 4 puntos, basada en el formato propuesto por Ventura-León (2019) y los propios criterios de la American Psychological Association, American Educational Research Association y National Council on Measurement in Education (APA, AERA, NCME, 2014). Las alternativas de respuesta han ido desde 1-Muy en desacuerdo a 4-Muy de acuerdo. Las categorías de análisis han sido: (1) relevancia, que tiene que ver con el grado de importancia del indicador con respecto al constructo; (2) representatividad, centrada en el grado de adecuación del indicador para representar al constructo; y (3) claridad, en cuanto al grado de precisión en la redacción del indicador, evitando posibles sesgos gramaticales.

Adicionalmente, en la plantilla de evaluación se ha incluido un apartado de observaciones y comentarios, tal y como proponen Escobar-Pérez y Cuervo-Martínez (2008).

## 2.5. Análisis de datos

En primer lugar, para cada categoría de análisis, el grado de acuerdo entre expertos y expertas se ha establecido mediante el índice de validez de contenido (del término en inglés content validity index; CVI) (Lynn, 1986; Martuza, 1977). Sobre este índice, Polit et al. (2007) señalan que el CVI muestra un desempeño altamente satisfactorio en la facilidad para el cálculo, la comprensión conceptual y la eficiencia en la obtención de información tanto a nivel de indicadores y/o ítems como de escala global.

Seguidamente, se ha calculado el índice de validez de contenido de un ítem (del término en inglés item-content validity index; I-CVI), que hace referencia a la proporción de expertos y expertas que califican el ítem con una puntuación de 3 o 4 (Lynn, 1986; Martuza, 1977; Polit et al., 2007).

Tras ello, a partir del promedio de los I-CVI, se ha calculado el índice de validez de contenido de la escala (del término en inglés scale-content validity index; S-CVI), que equivale a la cantidad relativa de ítems valorados con 3 o 4 puntos por la totalidad del panel de expertos y expertas (Polit et al., 2007).

Por último, con el objetivo de conseguir el acuerdo por azar, se ha aplicado el estadístico Kappa de Cohen modificado ( $k^*$ ), formulado por Polit et al. (2007), quienes indican que el valor de  $k^*$  puede interpretarse como regular/pasable, bueno o excelente según los estándares definidos por Cicchetti y Sparrow (1981) y Fleiss (1981).

Se ha creado una base de datos en el programa Microsoft Excel (versión 16.98), donde los datos han sido registrados y procesados estadísticamente. Se ha optado por este software debido a que, a diferencia de otros paquetes estadísticos, permite calcular el valor de  $k^*$  a través de distintas fórmulas combinadas.

## 2.6. Declaración ética

El protocolo de investigación de este estudio ha sido aprobado por el Comité de Ética de la Investigación Provincial de Huelva con código interno 0611-N-23. Se obtuvo previamente el consentimiento informado de los sujetos participantes en el panel de expertos y expertas, respetándose la normativa nacional e internacional aplicable (Reglamento General de Protección de Datos y Ley Orgánica 3/2018, de Protección de Datos Personales y garantía de los derechos digitales). Los datos fueron recopilados, almacenados y tratados de forma segura y anónima.

## 3. RESULTADOS

### 3.1. Valoración cuantitativa y análisis categórico global de indicadores

En la primera de las categorías de análisis (relevancia), los resultados del I-CVI han puesto de manifiesto que 87 de los 92 indicadores han obtenido valores de .80 a 1, siendo el porcentaje de acuerdo entre expertos y expertas sobre la relevancia de los indicadores del 80% al 100%. Concretamente, más del 76% de los indicadores han superado valores .90 del I-CVI. Asimismo,  $k^*$  ha oscilado de .799 a 1 en todos ellos. En cambio, 5 indicadores han presentado valores del I-CVI y  $k^* \leq .75$  (i2, i29, i72, i78 e i79) (tabla 1). Estos 5 indicadores se refieren al diseño y la accesibilidad visual (i2), al diseño instruccional y aspectos generales de la *app* (i29), y a elementos, recursos y estrategias de *gamificación*, así como la interactividad narrativa (i72, i78 e i79).

En cuanto al valor del S-CVI en esta primera categoría de relevancia, ha sido de .912.

**Tabla 1**

Resultados del I-CVI y  $k^*$  para la categoría de relevancia

Indicadores	I-CVI	$k^*$	%
72	.736	.73	1.08
2, 29, 78, 79	.75	.746	4.34
4, 13, 34, 39, 76	.80	.799	5.43
8, 20, 27, 44, 49, 51, 52, 56, 66, 81, 89, 90	.85	.849	13.04
6, 11, 12, 16, 17, 22, 23, 25, 28, 30, 42, 45, 48, 53, 54, 57, 60, 63, 68, 69, 71, 77, 80, 88	.90	.899	26.08
1, 7, 9, 15, 18, 19, 21, 31, 32, 35, 36, 38, 40, 41, 43, 46, 47, 50, 55, 59, 61, 62, 64, 65, 70, 74, 75, 82, 84, 86, 87, 91	.95	.949	34.78
3, 5, 10, 14, 24, 26, 33, 37, 58, 67, 73, 83, 85, 92	1	1	15.21

Nota. Elaborado a partir de los datos extraídos de Microsoft Excel (versión 16.98).

En la segunda de las categorías de análisis (representatividad), los resultados del I-CVI han arrojado que 88 de los 92 indicadores han alcanzado valores entre .80 y 1; es decir, el consenso sobre la representatividad de los indicadores se ha situado en un rango 80-100%, lográndose valores del I-CVI superiores a .90 en más del 72% de ellos. Además,  $k^*$  se ha mantenido dentro del rango .799-1 en este conjunto. En este caso, han sido 4 los indicadores con valores del I-CVI y  $k^* \leq .75$  (i2, i4, i16 e i72) (tabla 2). Estos 4 indicadores están relacionados con el diseño, la accesibilidad y la personalización visual (i2 e i4), la motivación y persuasión para el cambio de hábitos (i16), como también elementos, recursos y estrategias de *gamificación* (i72).

En esta segunda categoría de representatividad, el valor del S-CVI ha sido de .91.

**Tabla 2**

*Resultados del I-CVI y  $k^*$  para la categoría de representatividad*

Indicadores	I-CVI	$k^*$	%
72	.736	.73	1.08
2, 4, 16	.75	.746	3.26
17, 29, 79, 81	.80	.799	4.34
9, 11, 13, 27, 28, 39, 44, 46, 52, 54, 66, 69, 71, 76, 77, 78, 80	.85	.849	18.47
6, 8, 12, 22, 23, 25, 30, 34, 38, 40, 43, 45, 47, 50, 51, 53, 55, 60, 62, 63, 64, 65, 68, 74, 75, 88, 89, 90	.90	.899	30.43
1, 20, 24, 26, 32, 33, 37, 41, 42, 48, 49, 56, 57, 59, 61, 67, 73, 82, 83, 86, 87	.95	.949	22.82
3, 5, 7, 10, 14, 15, 18, 19, 21, 31, 35, 36, 58, 70, 84, 85, 91, 92	1	1	19.56

Nota. Elaborado a partir de los datos extraídos de Microsoft Excel (versión 16.98).

En la tercera y última de las categorías de análisis (claridad), los resultados del I-CVI han evidenciado que los 92 indicadores han obtenido valores de .80 a 1, por lo que el porcentaje de acuerdo entre expertos y expertas, en términos de claridad, ha sido de 80% a 100% en el listado completo, consiguiéndose valores del I-CVI >.90 en más del 86% de ellos. A su vez, el valor de  $k^*$  se ha situado entre .799 y 1 en la totalidad de indicadores (tabla 3).

Con respecto al valor del S-CVI en esta tercera categoría de claridad, este ha sido de .923.

**Tabla 3**

*Resultados del I-CVI y  $k^*$  para la categoría de claridad*

Indicadores	I-CVI	$k^*$	%
16, 81	.80	.799	2.17
72	.842	.841	1.08
8, 25, 53, 55, 61, 63, 74, 79, 82	.85	.849	9.78
1, 2, 6, 7, 13, 17, 23, 24, 29, 36, 38, 39, 40, 43, 46, 48, 50, 52, 54, 56, 57, 59, 60, 62, 64, 65, 66, 68, 69, 71, 73, 75, 76, 77, 78, 80, 84, 89	.90	.899	41.3
3, 11, 12, 14, 15, 20, 22, 26, 28, 30, 34, 35, 37, 41, 44, 45, 47, 49, 51, 58, 67, 70, 83, 86, 87, 88, 90	.95	.949	29.34
4, 5, 9, 10, 18, 19, 21, 27, 31, 32, 33, 42, 85, 91, 92	1	1	16.3
3, 5, 7, 10, 14, 15, 18, 19, 21, 31, 35, 36, 58, 70, 84, 85, 91, 92	1	1	19.56

Nota. Elaborado a partir de los datos extraídos de Microsoft Excel (versión 16.98).

### 3.2. Revisión de opiniones y sugerencias

Del panel de 20 expertos y expertas, 11 personas han realizado comentarios y/o sugerencias de modificación sobre el listado de indicadores.

La mayor parte de las observaciones se han centrado en la mejora de la redacción de los indicadores con el propósito de añadir información complementaria para aportar mayor claridad aún y facilitar su comprensión. Este ha sido el caso de los indicadores i1, i3, i5, i8, i13, i18, i23, i24, i26, i30, i36, i39, i43, i54, i66 e i83 (tabla 4).

**Tabla 4**

*Indicadores sugeridos para mejorar su redacción*

Indicadores con sugerencias de ajustes en la redacción	Contenido sintetizado del indicador	Información adicional recomendada
i1	Adecuación del diseño visual y funcional de la interfaz	Especificar niveles apropiados a cada perfil del público destinatario de la <i>app</i>
i3	Adecuación del tamaño y disposición de las fuentes	Especificar niveles apropiados a cada perfil del público destinatario de la <i>app</i>
i5	Adecuación de la estética visual	Los colores de la <i>app</i> tienen que ver con la legibilidad/contraste
i8	Eficiencia del consumo de recursos del dispositivo móvil	La expresión “no desperdicia recursos en exceso” es poco apropiada
i13	Evaluación de riesgo ante actualizaciones de terceros	Considerar posibles incidencias o interacciones
i18	Fluidez y transcurso lógico en la navegación	Añadir velocidad/fluidez
i23	Cumplimiento de normativas de accesibilidad	Detallar los estándares de accesibilidad necesarios atendiendo a la diversidad de usuarios y usuarias
i24	Accesibilidad lingüística del idioma	Incluir la posibilidad de cambiar el idioma de la <i>app</i>
i26	Coherencia entre la descripción oficial de la plataforma de descarga y las funcionalidades reales	Incluir al final del indicador la expresión “y se ajustan a la realidad”
i30	Indicación de no reemplazo de la atención profesional	Especificar área y/o especialidad de dichos profesionales
i36	Efectividad visual en la transmisión de conceptos	Hacer más hincapié en los aspectos visuales de la <i>app</i>
i39	Disponibilidad de función de búsqueda	Indicar el tipo de búsqueda (interna, externa, por palabras, etc.)
i43	Fundamentación ética de la información	Especificar principios y valores éticos según el tipo de sociedad o cultura
i54	Información sobre riesgos y efectos adversos por uso inadecuado	Indicar el tipo de riesgo potencial (personal, económico, de acceso de datos, etc.)
i66	Disponibilidad de mecanismos de seguridad	La expresión “gestión de contraseñas” no es suficientemente clara
i83	Beneficios para la salud derivados del uso	La <i>app</i> por sí misma no genera efectos beneficiosos en la salud, sino el conjunto de acciones derivadas de su uso

En línea con las sugerencias anteriores, los expertos y las expertas han coincidido en la necesidad de reformular los ya mencionados indicadores i29, i72 e i79 dado que su redacción resulta ambigua. Además, han manifestado que el indicador i16, anteriormente referido, podría no ser representativo de la dimensión en la que se encuentra (D3. Usabilidad de la *app*).

Asimismo, buena parte de las opiniones se han referido a posibles solapamientos entre indicadores. En este sentido, las propuestas de unificación han sido de i1 con i3; i2 con i4; i14 con i15; i22 con i23; i24 con i25; i32 con i33; i46 con i41 e i42; i48 con i53; i55 con i57; i70 con i75; e i82 con i84 (tabla 5).

**Tabla 5**

*Propuestas de unificación de indicadores*

Indicadores susceptibles de unificación	Contenido sintetizado del indicador	Indicadores con solapamiento conceptual
i1	Adecuación del diseño visual y funcional de la interfaz	i3: Si la disposición y el tamaño de las fuentes de la <i>app</i> son apropiados para el público destinatario
i2	Opciones de visualización del contenido	i4: Si la <i>app</i> permite cambiar el tamaño de las fuentes si fuera necesario
i14	Facilidad de uso	i15: Si es fácil aprender a usar la <i>app</i>
i22	Accesibilidad general y funcionalidades inclusivas	i23: Si la <i>app</i> cumple con los estándares reguladores de accesibilidad
i24	Accesibilidad lingüística del idioma	i25: Si la <i>app</i> sigue los principios de Diseño Universal (el público destinatario tiene acceso a todas las secciones y los recursos incluidos en la <i>app</i> independientemente de las capacidades individuales)
i32	Cantidad de información para lograr objetivos	i33: Si la información contenida en la <i>app</i> es exhaustiva y concisa (p. ej. enlaces a más información y recursos, no se incluyen datos irrelevantes ni explicaciones excesivas, etc.)
i46	Evidencia científica del contenido	i41: Si hay en la <i>app</i> referencias claras a las fuentes de información utilizadas (además de la autoría o el fabricante) i42: Si se indica explícitamente cuándo se crea la información que se utiliza y se reproduce en la <i>app</i>
i48	Información sobre políticas de privacidad	i53: Si la <i>app</i> advierte de si las actualizaciones pueden influir en datos sensibles (si cambia el uso de los datos o se recopilan datos diferentes)
i55	Protección y anonimización de datos confidenciales	i57: Si se garantiza la privacidad de la información registrada en la <i>app</i>
i70	Información y seguimiento del progreso	i75: Si la <i>app</i> incluye un ranking o una tabla de clasificación donde los jugadores y las jugadoras pueden ver su posición con respecto al progreso en el juego
i82	Efectividad o impacto en conocimientos y actitudes sobre hábitos saludables	i84: Si el usuario o la usuaria considera que la <i>app</i> le ha servido en el sentido de que ha habido una modificación en lo referente a hábitos saludables

Por otra parte, los expertos y las expertas han sugerido trasladar a otra dimensión los indicadores i9 e i13 (de la dimensión D2. Aspectos tecnológicos a la D3. Usabilidad de la *app*), ya que abordan cuestiones centradas en la usabilidad o seguridad en lugar de componentes técnicos y tecnológicos. También, han propuesto trasladar el indicador i68 de la dimensión D7. Gamificación a la D3. Usabilidad de la *app*, pues han expuesto que la integración de un tutorial se relaciona más con la experiencia de uso que con elementos y recursos lúdicos.

En cuanto a otras recomendaciones de mejora en el contenido de las dimensiones, los expertos y las expertas han argumentado que en la dimensión D4. Accesibilidad de la *app* no se han especificado los estándares actuales de accesibilidad. Más aún, en la dimensión D6. Protección de datos, han sugerido incorporar indicadores específicos sobre ciberseguridad y un indicador referido a la ubicación de datos almacenados en la nube debido a que es necesario conocer si se produce una transferencia de datos fuera de la Unión Europea. En esta misma dimensión, adicionalmente, han indicado que es importante detallar en su descripción que el diseño y el desarrollo de la *app*, así como los datos registrados en ella, deben estar amparados por el Reglamento General de Protección de Datos (RGPD) y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD). Por último, con respecto a la dimensión D8. Efectos del uso y resultados, han expresado que sería conveniente añadir un indicador centrado en experiencias docentes al utilizar la *app*.

#### 4. DISCUSIÓN Y CONCLUSIONES

Tras el proceso de validación a nivel cuantitativo y la consideración de las opiniones y sugerencias de los expertos y las expertas, finalmente se ha establecido un listado de 72 indicadores, los cuales son susceptibles de ser materializados en un instrumento integral de evaluación de la calidad de *apps* educativas y gamificadas destinadas a la promoción de hábitos de vida saludable en la población infantil y adolescente. Puede acceder a dicho listado nuevamente a través de (Delgado-Morales y Duarte-Hueros, 2025).

Los indicadores que se han examinado en profundidad, para ser modificados, reubicados en otra dimensión y/o eliminados, a partir de los índices y estadísticos de la valoración cuantitativa para cada categoría de análisis, han sido i2, i4, i16, i29, i72, i78 e i79.

En este sentido, según Lynn (1986) y Polit et al. (2007), en concordancia con el número de expertos y expertas que han conformado el panel, los indicadores con valores no aceptables del I-CVI ( $I-CVI < .78$ ) en la categoría de relevancia, han sido i2, i29, i72, i78 e i79. En el caso de la categoría de representatividad, han sido de nuevo los indicadores i2 e i72, además de i4 e i16. En cambio, todos los indicadores en la categoría de claridad han arrojado valores del I-CVI bastante satisfactorios puesto que no se han dado índices inferiores a  $.78$  (Lynn, 1986; Polit et al., 2007).

Acerca de los valores del S-CVI, en consonancia con Lynn (1986), se han considerado como excelentes al ser todos ellos superiores a  $.90$  (S-CVI relevancia= $.912$ ; S-CVI representatividad= $.91$  y S-CVI claridad= $.923$ ).

Referente a los valores de  $k^*$  en las tres categorías de análisis evaluadas, conforme a los criterios de Polit et al. (2007), los indicadores en su mayoría han mostrado resultados excelentes ( $k^* = .799-1$ ), situándose todos ellos en el intervalo bueno-excelente y guardando alta relación con el constructo subyacente.

Si bien la mayor parte de los ajustes en los indicadores se han realizado en base a los resultados cuantitativos, con una concordancia interpretada como casi perfecta, y es sabido que analizar las características de los ítems

es fundamental para determinar la precisión de la medición de un instrumento (Bernal-García et al., 2020), las sugerencias y recomendaciones de los expertos y las expertas, coincidiendo con Cabello-Ochoa y Carrera-Farran (2017), han sido claves en la configuración definitiva del listado de indicadores, aportando información de gran valor en la toma de decisiones.

En definitiva, en términos generales, las valoraciones de los expertos y las expertas han sido sumamente positivas, destacando que los indicadores planteados podrían resultar muy útiles para orientar a profesionales implicados en las fases de diseño y desarrollo de este tipo de software educativo y otros recursos digitales, así como posibilitar una evaluación tecnológica global a través de la construcción y mejora continua de instrumentos versionados y adaptados que evalúen la calidad de la tecnología educativa en diferentes contextos de aplicación. No obstante, habrá que considerar el obstáculo estructural que supone la obsolescencia programada de la tecnología en cuanto a la continuidad de uso y la actualización sostenible en el tiempo de las *apps* en cuestión, todo ello para ofrecer soluciones educativas válidas y fiables ajustadas a un entorno cambiante.

En relación a la selección de las personas para el juicio experto, en este trabajo se ha optado por criterios cualitativos ampliamente aceptados en la literatura, priorizando la diversidad disciplinar y la experiencia profesional como elementos clave para garantizar la validez de contenido, si bien, la incorporación de coeficientes o indicadores cuantitativos como el Coeficiente de Competencia Experta (Kc), el Coeficiente de Conocimiento (K) y el Coeficiente de Argumentación (Ka) podrían aportar mayor rigor metodológico. Se recomienda su inclusión en futuras investigaciones.

## 5. FINANCIACIÓN

Este trabajo ha sido apoyado por fondos públicos y en concurrencia competitiva por la Estrategia de Política de Investigación y Transferencia de la Universidad de Huelva (ayudas predoctorales de Personal Investigador en Formación, 2020).

## 6. CONTRIBUCIÓN DE LOS AUTORES

Conceptualización, C.D.M. y A.D.H.; curación de datos, C.D.M.; análisis formal, C.D.M.; investigación, C.D.M. y A.D.H.; metodología, C.D.M. y A.D.H.; software, C.D.M.; supervisión, A.D.H.; validación, C.D.M. y A.D.H.; redacción—preparación del borrador original, C.D.M. y A.D.H.; redacción—revisión y edición, C.D.M. y A.D.H.

## 7. REFERENCIAS

Alexandrou, C., Henriksson, H., Henström, M., Henriksson, P., Nyström, C. D, Bendtsen, M. & Löf, M. (2023). Effectiveness of a Smartphone App (MINISTOP 2.0) integrated in primary child health care to promote healthy diet and physical activity behaviors and prevent obesity in preschool-aged children: randomized controlled trial. *The International Journal of Behavioral Nutrition and Physical Activity*, 20(1), 22. <https://doi.org/10.1186/s12966-023-01405-5>

American Educational Research Association, American Psychological Association y National Council on Measurement in Education. (2014). *The Standards for Educational and Psychological Testing*. American Educational Research Association.

- Ayre, C. & Scally, A. J. (2014). Critical Values for Lawshe's Content Validity Ratio: Revisiting the Original Methods of Calculation. *Measurement and Evaluation in Counseling and Development*, 47(1), 79-86. <https://doi.org/10.1177/0748175613513808>
- Azad-Khaneghah, P., Neubauer, N., Cruz, A. M. & Liu, L. (2021). Mobile health app usability and quality rating scales: a systematic review. *Disability and Rehabilitation: Assistive Technology*, 16(7), 712-721. <https://doi.org/10.1080/17483107.2019.1701103>
- Bernal-García, M. I., Salamanca-Jiménez, D. R., Perez-Gutiérrez, N. y Quemba-Mesa, M. P. (2020). Validez de contenido por juicio de expertos de un instrumento para medir percepciones físico-emocionales en la práctica de disección anatómica. *Educación Médica*, 21(6), 349-356. <https://doi.org/10.1016/j.edumed.2018.08.008>
- Brown, J. M., Franco-Arellano, B., Froome, H., Siddiqi, A., Mahmood, A. & Arcand, J. (2022). The Content, Quality, and Behavior Change Techniques in Nutrition-Themed Mobile Apps for Children in Canada: App Review and Evaluation Study. *JMIR mHealth and uHealth*, 10(2), e31537. <https://doi.org/10.2196/31537>
- Cabello-Ochoa, S. y Carrera-Farran, F. X. (2017). Diseño y validación de un cuestionario para conocer las actitudes y creencias del profesorado de educación infantil y primaria sobre la introducción de la robótica educativa en el aula. *EduTec, Revista Electrónica De Tecnología Educativa*, (60), a362. <https://doi.org/10.21556/edutec.2017.60.871>
- Cheng, H., Tutt, A., Llewellyn, C., Size, D., Jones, J., Taki, S., Rossiter, C. & Denney-Wilson, E. (2020). Content and Quality of Infant Feeding Smartphone Apps: Five-Year Update on a Systematic Search and Evaluation. *JMIR mHealth and uHealth*, 8(5), e17300. <https://doi.org/10.2196/17300>
- Cicchetti, D. V. & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127-137.
- Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, 5(4), 194-197. [https://doi.org/10.1016/s0897-1897\(05\)80008-4](https://doi.org/10.1016/s0897-1897(05)80008-4)
- Delgado-Morales, C. y Duarte-Hueros, A. (2025). Validez de contenido de indicadores para evaluar la calidad de aplicaciones móviles dirigidas a menores que promueven la educación en salud a través de la gamificación [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.15870745>
- Delgado-Morales, C. y Duarte-Hueros, A. (2023). Una Revisión sistemática de instrumentos que evalúan la calidad de aplicaciones móviles de salud. Pixel-Bit. *Revista De Medios Y Educación*, 67, 35-58. <https://doi.org/10.12795/pixelbit.97867>
- Escobar-Pérez, J. y Cuervo-Martínez, Á. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6(1), 27-36.
- Expósito-López, J., Romero-Díaz de la Guardia, J. J. y 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>

- Fijačko, N., Gosak, L., Cilar, L., Novšak, A., Creber, R. M., Skok, P. & Štiglic, G. (2020). The Effects of Gamification and Oral Self-Care on Oral Hygiene in Children: Systematic Search in App Stores and Evaluation of Apps. *JMIR mHealth and uHealth*, 8(7), e16365. <https://doi.org/10.2196/16365>
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.). John Wiley & Sons.
- Frontini, R., Sousa, P., Dixe, M. A., Ferreira, R. & Figueiredo, M. C. (2020). Designing a mobile app to promote healthy behaviors and prevent obesity: analysis of adolescents' preferences. *Informatics for Health and Social Care*, 45(3), 327-341. <https://doi.org/10.1080/17538157.2020.1725766>
- Ho, T. C., McGrath, C., Yiu, C. K. & Lee, G. H. (2022). Apps for Promoting Children's Oral Health: Systematic Search in App Stores and Quality Evaluation. *JMIR Pediatrics and Parenting*, 5(2), e28238. <https://doi.org/10.2196/28238>
- Hyrkäs, K., Appelqvist-Schmidlechner, K. & Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies*, 40(6), 619-625. [https://doi.org/10.1016/s0020-7489\(03\)00036-1](https://doi.org/10.1016/s0020-7489(03)00036-1)
- Instituto Nacional de Estadística. (2024). *Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares 2024*. <https://bit.ly/3HbmcWe>
- Jeminiwa, R. N., Hohmann, N. S. & Fox, B. I. (2019). Developing a Theoretical Framework for Evaluating the Quality of mHealth Apps for Adolescent Users: A Systematic Review. *The Journal of Pediatric Pharmacology and Therapeutics*, 24(4), 254-269. <https://doi.org/10.5863/1551-6776-24.4.254>
- Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382-386.
- Martuza, V. R. (1977). *Applying Norm-Referenced and Criterion-Referenced Measurement in Education*. Allyn and Bacon.
- McMillan, J. H. y Schumacher, S. (2005). *Investigación educativa. Una introducción conceptual* (5.ª ed.). Pearson Educación.
- Merino-Godoy, M. Á., Yot-Domínguez, C., Conde-Jiménez, J. & de la Calle-Cabrera, A. M. (2022). Confirmatory Validation of an Evaluation Instrument for Interventions Based on the Healthy Habits App Healthy Jeart with Adolescents. *Journal of Personalized Medicine*, 12(3), 470. <https://doi.org/10.3390/jpm12030470>
- Palacios-Gálvez, M. S., Yot-Domínguez, C. y Merino-Godoy, Á. (2020). Healthy Jeart: promoción de la salud en la adolescencia a través de dispositivos móviles. *Revista Española de Salud Pública*, 94, e202003010. <https://doi.org/10.4321/S1135-57272020000100008>
- Polit, D. F., Beck, C. T. & Owen, S. V. (2007). Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations. *Research in Nursing & Health*, 30(4), 459-467. <https://doi.org/10.1002/nur.20199>
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S. & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. <https://doi.org/10.1093/swr/27.2.94>

- Sefa-Yeboah, S. M., Osei-Annor, K., Koomson, V. J., Saalia, F. K., Steiner-Asiedu, M. & Mills, G. A. (2021). Development of a Mobile Application Platform for Self- Management of Obesity Using Artificial Intelligence Techniques. *International Journal of Telemedicine and Applications*, 2021, 6624057. <https://doi.org/10.1155/2021/6624057>
- Skjong, R. & Wentworth, B. H. (2001, June 17-22). *Expert Judgement and Risk Perception* [Conference paper]. The Eleventh International Offshore and Polar Engineering Conference, Stavanger, Norway. <https://bit.ly/3kZayAM>
- Sousa, P., Martinho, R., Reis, C. I., Dias, S. S., Gaspar, P. J., Dixe, M. A., Luis, L. S. & Ferreira, R. (2020). Controlled trial of an mHealth intervention to promote healthy behaviours in adolescence (TeenPower): Effectiveness analysis. *Journal of Advanced Nursing*, 76(4), 1057-1068. <https://doi.org/10.1111/jan.14301>
- Tarricone, R., Petracca, F., Cucciniello, M. & Ciani, O. (2022). Recommendations for developing a lifecycle, multi-dimensional assessment framework for mobile medical apps. *Health Economics*, 31(S1), 73-97. <https://doi.org/10.1002/hec.4505>
- Thornton, L., Gardner, L. A., Osman, B., Green, O., Champion, K. E., Bryant, Z., Teesson, M., Kay-Lambkin, F. & Chapman, C. (2021). A Multiple Health Behavior Change, Self-Monitoring Mobile App for Adolescents: Development and Usability Study of the Health4Life App. *JMIR Formative Research*, 5(4), e25513. <https://doi.org/10.2196/25513>
- Ventura-León, J. (2019). De regreso a la validez basada en el contenido. *Adicciones*. <https://doi.org/10.20882/adicciones.1213>
- Violini, M. L., Sanz, C. V. y Pesado, P. M. (2022). Prototipo de un framework para la creación de Objetos de Aprendizaje y su evaluación mediante juicio de expertos. *EduTec, Revista Electrónica De Tecnología Educativa*, (81), 137-154. <https://doi.org/10.21556/edutec.2022.81.2529>
- Witkin, B. R. & Altschuld, J. W. (1995). *Planning and Conducting Needs Assessment: A Practical Guide*. Sage Publications.
- World Health Organization. (2011). *mHealth: new horizons for health through mobile technologies: second global survey on eHealth*. Global Observatory for eHealth. <https://bit.ly/3LV8iVu>

#### Para citar este artículo:

Delgado-Morales, C., y Duarte-Hueros, A. (2026). Validez de contenido de indicadores para evaluar la calidad de aplicaciones móviles dirigidas a menores que promueven la educación en salud a través de la gamificación. *EduTec, Revista Electrónica de Tecnología Educativa*, (95), Art. 9. <https://doi.org/10.21556/edutec.2026.95.4347>