

Métricas que podrían usarse en el Desarrollo de Aplicaciones Móviles para Personas con deficiencias Visuales: Una Revisión de Literatura Preliminar

Erick Zambrano¹, Rolando P. Reyes Ch.¹, John W. Castro², Efraín R. Fonseca C.¹

ejzambrano, rpreyes1, erfonseca@espe.edu.ec, john.castro@uda.cl

¹ Universidad de las Fuerzas Armadas ESPE, Departamento de Ciencias de la Computación, Av. General Rumiñahui s/n y Ambato, EC171103, Sangolquí, Ecuador.

² Universidad de Atacama, Departamento de Ingeniería Informática y Ciencias de la Computación, Av. Copayapu 485, 1530000, Copiapó, Chile.

Pages: 985–999

Resumen: Antecedentes: La masificación del uso de dispositivos móviles ha propiciado el desarrollo paralelo de un sinnúmero de aplicaciones móviles (apps), cuyas características son explotadas por la mayoría de las personas. No obstante, existen segmentos de personas con deficiencias visuales (PDV) que han sido marginadas por temas de usabilidad y/o accesibilidad. **Objetivo:** Identificar estudios primarios que reporten métricas de usabilidad y/o accesibilidad aplicable al desarrollo y evaluación de sistemas móviles, incluyendo las necesidades de las PDV. **Método:** Se llevó a cabo una Revisión Sistemática de Literatura Preliminar (PLR) basada en las directrices de una Revisión Sistemática de Literatura (SLR). **Resultados:** Se obtienen 6 estudios primarios que reportan 73 métricas para apps, las cuales fueron agrupadas en 8 contextos de usabilidad: Contenido visual, eficacia, navegabilidad, eficiencia, aprendizaje, satisfacción, comprensión y accesibilidad. **Conclusión:** Los contextos de mayor relevancia son: Contenido visual, eficacia, navegabilidad, comprensión y accesibilidad, siendo los 3 primeros los que agrupan la mayor cantidad de métricas de usabilidad.

Palabras-clave: Usabilidad; Accesibilidad; Métricas; Ingeniería de Software; Revisión de Literatura

Metrics That Could Be Used In Mobile Applications Development For People With Visual Impairments: A Preliminary Literature Review

Abstract: Background: The widespread use of mobile devices has motivated the parallel development of countless mobile applications (apps), whose characteristics benefit most people. However, there are segments of people with visual impairments (PVI) who have been marginalized due to usability and accessibility issues. **Aim:** To identify primary studies which report metrics of usability, accessibility or both, applicable to the development and assessing of mobile systems, including requirements of PVI. **Method:** A Preliminary Literature Review (PLR) was done

based on the guidelines of a Systematic Literature Review (SLR). **Results:** 6 primary studies reported 73 metrics for apps, which were grouped into 8 usability contexts: visual content, effectiveness, navigability, efficiency, learnability, satisfaction, understandability and accessibility. **Conclusions:** The most relevant contexts are Visual content, effectiveness, navigability, understandability, and accessibility; of which, the first three group the highest number of usability metrics.

Keywords: Usability; Accessibility; Metrics; Software Engineering; Literature Review.

1. Introducción

La masificación del uso de dispositivos móviles para realizar tareas cotidianas desde cualquier lugar a través de internet ha crecido exponencialmente en los últimos años (Dias Silva, Carvalho Nepomuceno, Silva, & Freire Pimenta, 2018). De hecho, se espera que en 2019 la cantidad de usuarios de teléfonos móviles a nivel mundial supere los cinco mil millones, lo que representa aproximadamente el 67 % de la población mundial (Baron, 2018). Con este fin, se han desarrollado paralelamente un sinnúmero de aplicaciones móviles (apps), las cuales son ampliamente utilizadas por la mayoría de los usuarios. No obstante, segmentos de usuarios con deficiencias físicas han sido marginados, dado que enfrentan dificultades al utilizar los sistemas móviles, especialmente las personas con deficiencias visuales (PDV) (Dias Silva, Carvalho Nepomuceno, Silva, & Freire Pimenta, 2018).

Las PDV enfrentan una serie de retos al momento de interactuar con un sistema móvil, principalmente relacionados con la usabilidad y accesibilidad referidas al software y hardware respectivamente (Carvalho & Freire, 2017). En lo que respecta al software, está comprobado que, durante el desarrollo de la mayoría de las apps, no son considerados criterios ni parámetros referentes a la usabilidad para las PDV; mientras que, en lo relacionado con el hardware, los dispositivos móviles presentan una gran variedad de características (tamaños variados, colores diversos, etc.), siendo la más común el uso de pantallas táctiles como medio de interacción (TIC, 2013). Esta situación dificulta e incluso imposibilita el uso de ciertas apps; así como también, reduce significativamente la interacción PDV - sistema móvil (Kane, 2009).

La usabilidad es un atributo de calidad muy importante dentro del desarrollo de software; de hecho, existe una innumerable cantidad de estudios al respecto (P. Di , y otros, 2012). No obstante, los estudios de usabilidad enfocados en las apps han llamado la atención recientemente, al parecer por el incremento en su uso. Para el caso de las PDV, la importancia de la usabilidad se acrecienta aún más, por obvias razones. Sin embargo, el número y la madurez de los estudios al respecto nos indican que, al parecer, no se está considerando a las PDV al momento de construir y evaluar la usabilidad de las apps, limitando así la inclusión de este segmento de la población en la sociedad (Facanha Rocha, Cavalcante, & Campos, 2014).

Esta situación ha puesto de manifiesto la necesidad de identificar trabajos en los que se estudien métricas de usabilidad y accesibilidad que podrían ser aplicadas en el desarrollo y evaluación de los sistemas móviles, incluyendo las necesidades de las PDV. Como estrategia metodológica para obtener los estudios requeridos, consideramos

la realización de una Revisión de Literatura Preliminar (PLR de las siglas del inglés Preliminary Literature Review). La propuesta de un estudio preliminar responde a dos situaciones: por una parte, a optimizar el tiempo que normalmente es utilizado en una Revisión Sistemática de Literatura (SLR de las siglas del inglés Systematic Literature Review); y por otra parte, la necesidad de obtener de manera rigurosa estudios fiables. Como resultado, se obtuvieron 6 estudios primarios en los que se hacen referencia a contextos de métricas de: contenido visual (por ejemplo, visual coherence), eficacia (por ejemplo, ease of use), navegabilidad (por ejemplo, display self adjustment), satisfacción (por ejemplo, reach goal accurately), comprensión (por ejemplo, display load) y accesibilidad (por ejemplo, information access easily). Estos contextos de métricas fueron considerados idóneos para evaluar la usabilidad de apps que podrían ser utilizadas por las PDV, dada su afinidad con las características de los diferentes tipos de deficiencias visuales.

2. Antecedentes

La usabilidad es considerada como una característica de calidad muy importante dentro de la Ingeniería de Software (IS). Esta característica normalmente suele ser considerada durante el desarrollo de las aplicaciones de escritorio y web (Carvajal, Moreno, & M. Segura, 2013). No obstante, en los últimos años el desarrollo se ha orientado hacia las apps, debido al incremento exponencial en el uso de los dispositivos móviles (Joorabchi, Mesbah, & Kruchten, 2013). Ahora bien, evaluar la usabilidad de las apps, no es una tarea fácil. Los ingenieros de software necesitan conocer exactamente qué es lo que hay que medir, a partir de un sinnúmero de dominios que han sido propuestos en la literatura (Baharuddin, Singh, & Razali, 2013). Razón por la cual, es frecuente la selección errónea de métricas de usabilidad para las apps, dado que inicialmente las métricas fueron definidas para aplicaciones de escritorio o web. El problema radica en que los sistemas móviles presentan características particulares como: dimensiones reducidas, ancho de banda limitado, tamaño de la pantalla, entre otros (Hanlu, Malu, & Findla, 2014).

La literatura que trata la usabilidad es muy extensa. Hemos clasificado la literatura relacionada con métricas de usabilidad en dos grupos. En el primero, se encuentran los trabajos que identifican métricas para aplicaciones de escritorio o web, como el estándar ISO/IEC 9126 que define métricas de usabilidad. Sagar and Saha (Anju & Sagar, 2017) realizan una SLR para identificar características de usabilidad generales (por ejemplo: estándares, métodos, métricas, dominios de usabilidad). Los autores identificaron 150 estudios primarios, a partir de los cuales concluyen que no existe un modelo consensuado para evaluar la usabilidad en diferentes dominios. El trabajo de Sauro y Lewis (Sauro & Lewis R., 2009) reporta factores que pueden incidir en la correlación entre distintas métricas de usabilidad, a partir de una base de información constituida de 90 estudios de usabilidad.

El segundo grupo está relacionado con la evaluación de la usabilidad de las apps. Ahmad & otros (Ahmad, Rextin, & Kulsoom, 2018) logran sintetizar una lista de guías relacionadas con la usabilidad para el desarrollo de aplicaciones multiplataforma en dispositivos móviles. Los autores obtienen 148 estudios primarios, que proponen un total

de 359 guías que fueron organizadas en 7 categorías. Además, aplican un experimento controlado con estudiantes de grado y posgrado con el objetivo de evaluar algunas apps y detectar problemas de usabilidad. Kumar y Mohite (Kumar & Mohite, 2018) realizan una SLR en la que identifican 23 estudios primarios relacionados con la evaluación de la usabilidad de apps para el aprendizaje. Lettner y Holzmann (Lettner & Holzmann, 2012) proponen un nuevo enfoque, de la mano de un conjunto de herramientas para la evaluación automatizada y no supervisada de apps. Estas herramientas permiten identificar y visualizar fallos en el diseño (por ejemplo, errores en la navegación o eficiencia). Kronbauer & otros (Kronbauer H & Santos S., 2012) proponen un modelo de evaluación de usabilidad híbrido para apps, el cual describe cómo supervisar y recopilar automáticamente datos de contexto y métricas de usabilidad.

En resumen, podemos indicar que existen trabajos relacionados con las métricas de usabilidad para aplicaciones de escritorio o web y otros relacionados con las métricas de usabilidad para apps. No obstante, no encontramos estudios sobre métricas de usabilidad para apps que consideren las necesidades de las PDV, lo cual nos pareció extraño, dado que según la Organización Mundial de la Salud (OMS) (World Health Organization, 2017), actualmente existen 314 millones de personas en el mundo que tienen discapacidades visuales y que por lo antes mencionado son excluidos del uso de la tecnología (Kulpa Costa & Amaral Gon). Por tanto, se hace necesario investigar que métricas de accesibilidad o usabilidad podrían ser aplicadas durante el desarrollo de apps para las PDV. Para tal efecto, nos hemos planteado la siguiente pregunta de investigación (RQ):

RQ: ¿Qué métricas de accesibilidad y/o usabilidad, consideradas comúnmente en la construcción del software, podrían ser utilizadas para evaluar las apps desde la perspectiva de las PDV?

3. Método de Investigación

Para dar respuesta a nuestra RQ realizamos una PLR como estrategia metodológica para la cual consideramos utilizar algunas fases, etapas y actividades propuestas por Kitchenham & otros (Kitchenham, Brereton, O , & Linkman, 2009) para realizar una SLR, dada su representatividad en este tipo de estudios en la IS. Más específicamente, consideramos: planificar la revisión y ejecutar la revisión. De la fase planificar la revisión consideramos las etapas: (1) identificar la necesidad de una revisión, (2) especificar la pregunta de investigación y (3) desarrollar el protocolo de investigación. Estas etapas permitieron orientar nuestra PLR. Para la fase ejecución de la revisión seleccionamos las etapas: (1) identificación de la investigación, (2) selección de estudios, (3) evaluación de calidad de los estudios y (4) extracción de datos y síntesis. Esta última etapa es una fusión de las etapas extracción de datos y síntesis de datos. Esta fusión responde a que posiblemente los datos resultantes no son extensos y por tanto, permitan una síntesis mayormente descriptiva (Velásquez, 2015).

3.1. Desarrollo del Protocolo e Identificación de la Investigación

Las actividades realizadas en esta etapa son: (1) Conformación del grupo de control, (2) identificación y selección de las palabras claves, (3) especificación de los criterios

de inclusión y exclusión y (4) conformación de la cadena de búsqueda. A continuación, describimos cada una de estas etapas.

Conformación del Grupo de Control: Según Zhang & otros (Zhang, Ali Babar, & Tellb, 2010), un grupo de control es un conjunto de estudios relevantes que permiten establecer otros estudios similares bajo criterios explícitos establecidos en las RQ. Desde esta perspectiva, nuestro grupo de control de literatura nos permitió identificar y seleccionar las palabras comúnmente utilizadas en el contexto de la temática de investigación, las cuales constituyeron la base para conformar de la cadena de búsqueda. Así mismo, el grupo de control nos permitió afinar los criterios de inclusión y exclusión que nos habíamos planteado inicialmente para definir las características de los estudios que buscamos. Para la conformación del grupo de control son necesarios 4 pasos (Zhang, Ali Babar, & Tellb, 2010).

- Identificación de Investigadores. Contamos con un investigador novato y un investigador experto con una amplia experiencia en la ejecución de revisiones sistemáticas de literatura. Estos investigadores se encargaron de conformar el grupo de control.
- Búsqueda de Artículos Candidatos para el Grupo de Control. De forma individual, los investigadores buscan artículos aparentemente idóneos para el grupo de control. Esta búsqueda fue realizada en bases de datos digitales tales como Scopus, ACM, IEEE, entre otras (Kitchenham, Brereton, O, & Linkman, 2009). Posteriormente, los investigadores examinaron el título, resumen y palabras claves, guiados por los criterios de inclusión y exclusión preliminares. Como resultado, un total de 27 artículos fueron propuestos para la conformación del grupo de control (17 por el investigador novato y 10 por el investigador experto).
- Validación y Conformación del Grupo de Control. Una vez obtenidos los artículos candidatos para conformar el Grupo de Control, se realiza una validación cruzada entre los investigadores. Esta validación consiste en la revisión del texto completo de cada uno de los 27 artículos candidatos, con el propósito de verificar que los artículos tengan una relación directa con la RQ. Como resultado de la validación cruzada se conformó el Grupo de Control con los siguientes artículos: (Campanella, 2012), (Ba Matrafi & Hussain, 2018), (Molano I., Yara S. , & García K. , 2015), (FortesPontin , Salgado, de Souza Santos, Agostini do Amaral, & Nogueira da Silva , 2017), (Cruz Reis, Gomes, Dias-Neto, & de L. Fontão, 2015), (Baumel, Faber, Mathur, Kane, & Muench, 2017), (Sánchez Sierra & Togores, 2012), (Guerrero & Vega, 2018), (Zali & Ahmad Fadzliah , 2016), (Kumanan , y otros, 2016), (Olmedo-Moreno & López-Delgado , 2015). En <https://goo.gl/nmYxZw> puede consultarse el detalle de la validación cruzada y la selección del grupo de control.

Identificación y Selección de las Palabras Claves: En los artículos del grupo de control identificamos las palabras que aparecen con mayor frecuencia (Bolderston, 2008) con la ayuda del software ATLAS.ti¹, lo cual puede ser revisado en detalle en². La **Tabla 1** presenta, para cada una de las 10 palabras con mayor frecuencia, el número de veces que es usada en el grupo de control y la cobertura que consiste en el porcentaje

¹ <https://atlasti.com/>

² <https://goo.gl/nmYxZw>

de artículos del grupo de control donde aparece la palabra. Las palabras seleccionadas fueron consideradas para conformar y afinar la cadena de búsqueda.

Palabras	Conteo Cobertura Palabras		Conteo Cobertura	
usability	451	90 %	quality	149 64 %
mobile	378	90 %	applications	130 72 %
evaluation	227	90 %	application	118 55 %
accessibility	191	72 %	model	103 55 %
app	154	36 %	metrics	74 45 %

Tabla 1 – Palabras con Mayor Frecuencia en el Grupo de Control

Conformación de la Cadena de Búsqueda: Con base en las palabras seleccionadas se construyó la cadena de búsqueda luego de varios ciclos de afinamiento. Para la construcción de la cadena de búsqueda se utilizó el operador lógico AND para incluir palabras que pertenecen a contextos diferentes y el operador OR para incluir sinónimos de palabras que pertenecen al mismo contexto. El afinamiento consiste en probar distintas cadenas variando su contenido de palabras a modo prueba-error, con el propósito de obtener:

- Un número razonable de estudios (cientos no miles)
- Un alto porcentaje de estudios que se relacionen directamente con la temática de investigación
- La mayoría de los estudios del Grupo de Control dentro de la respuesta de la cadena
- En el proceso de afinamiento se generaron un total de 6 cadenas de búsqueda hasta encontrar la cadena ideal. Este proceso se realizó en la base digital Scopus. Para validar la cadena de búsqueda más adecuada, se utilizaron diferentes ratios:
- Ratio-X (ERGC/TER): Orientado a establecer el ratio de los artículos positivos - positivos recuperados del total de artículos encontrados con la cadena de búsqueda. Este ratio se calcula dividiendo la cantidad de estudios recuperados del grupo de control (ERGC) con la cadena X sobre el número total de estudios recuperados con la cadena X (TER).
- Ratio-Y (ERGC/GC): Establece el ratio de los artículos del grupo de control recuperados. Este ratio se calcula dividiendo la cantidad de estudios recuperados del grupo de control (ERGC) con la cadena X sobre el número total de estudios del Grupo de Control (GC).

Con base en el cálculo de estos ratios, se logró determinar que la cadena C6 (Ratio-X = 0.09, Ratio-Y = 0.64) es la más adecuada para encontrar los estudios primarios. Las cadenas afinadas (C1-C5) y la cadena C6 se muestran en la **Tabla 2**.

Especificación de los Criterios de Inclusión y Exclusión: A partir de las características y particularidades de los artículos del grupo de control, los investigadores afinaron los criterios de inclusión y exclusión. Los criterios de inclusión corresponden a artículos que:

	Search Strings	Recov.	Ratio-X	Cont. Group	Ratio-Y
		studies			
C1	usability AND metrics AND software AND quality AND model	88	0.02	[25], [30]	0.18
C2	usability AND metrics AND software AND quality AND mobile AND apps	3	0.67	[25], [32]	0.18
C3	usability AND metrics AND mobile AND (applications OR app OR apps) AND design AND development	17	0.12	[27], [28]	0.18
C4	(accessibility OR usability) AND metrics AND mobile AND (apps OR app OR applications) AND design AND development	19	0.10	[27], [28]	0.18
C5	accessibility AND metrics AND mobile	53	0.05	[23], [24], [26]	0.27
C6	(accessibility OR usability) AND metrics AND evaluation AND mobile AND (apps OR app OR applications)	78	0.09	[23], [24], [25], [27], [28], [31], [32]	0.64

Tabla 2 – Cadenas de Búsqueda Piloteadas y sus Ratios Asociados

- Describen métricas de accesibilidad o usabilidad utilizadas en el desarrollo de apps.
- Proponen métricas de usabilidad para el desarrollo de apps.
- Reportan métricas de usabilidad para personas con discapacidad.
- Han sido publicados en revistas, congresos, o en capítulos de libros (indexados por lo menos en Scopus).
- Los criterios de exclusión son los siguientes:
- Artículos que no describen o no presenten información sobre métricas de accesibilidad o usabilidad utilizadas durante el desarrollo de apps.
- Artículos que reporten casos de estudio o metodologías para el diseño de apps.
- Artículos escritos en idiomas diferentes al inglés.

3.2. Selección de Estudios

La búsqueda de estudios fue realizada en las bases digitales IEEE Xplore³, Scopus⁴ y ACM Digital Library⁵ usando la cadena **C6**. En una primera instancia, logramos encontrar un total de 108 estudios candidatos distribuidos de la siguiente manera: 78 estudios candidatos en Scopus, 3 estudios en IEEE Xplore y 27 estudios en ACM Digital Library. Posteriormente, eliminamos estudios duplicados entre las bases de datos. De esta depuración se descarta 1 estudio candidato de IEEE Xplore y 10 estudios candidatos de ACM Digital Library. Como resultado de esta primera depuración se obtienen 97 estudios candidatos.

³ <http://ieeexplore.ieee.org/>

⁴ <https://www.scopus.com/>

⁵ <https://dl.acm.org/>

A paso seguido realizamos (investigadores novato y experto) un primer filtro aplicando los criterios de inclusión y exclusión sobre el título, resumen y palabras claves de cada uno de los 97 estudios candidatos. Como resultado de este primer filtro se obtuvieron 14 estudios (83 estudios candidatos fueron descartados). Los estudios obtenidos a partir del primer filtro fueron evaluados nuevamente en un segundo filtro, lo cual se describe a continuación.

3.3. Evaluación de Calidad de los Estudios

Los 14 estudios obtenidos a partir del primer filtro fueron sometidos a un segundo filtro. En este filtro interviene un tercer investigador, el cual, participa con fines de rigurosidad en la validación cruzada. Previo a la validación, cada investigador aplica los criterios de inclusión y exclusión al texto completo de cada uno de los 14 estudios.

Como resultado del segundo filtro se obtienen 6 estudios primarios (ver **Tabla 3**), que hemos codificado como “SP#”. En el enlace <https://goo.gl/nmYxZw> se encuentra el detalle del proceso de selección de estudios primarios.

Cod.	Año	Título	Ref.
SP1	2001	QUIM: A Framework for Quantifying Usability Metrics in Soft-ware Quality Models	[35]
SP2	2012	Mobile Learning New forms of education	[36]
SP3	2016	An initial theoretical usability evaluation model for assessing defence mobile e-based application system	[31]
SP4	2018	Modeling Measurement Metrics for E-Book App on Mobile Devices	[24]
SP5	2017	Enlight: A Comprehensive Quality and Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions	[28]
SP6	2018	Usability analysis: Is our software inclusive?	[30]

Tabla 3 – Estudios Primarios

4. Resultados e Interpretación

Esta sección presenta los resultados obtenidos en la PLR, los cuales están divididos en dos partes: Generalidades de los estudios primarios y la respuesta a la pregunta de investigación.

La primera parte se encuentra resumida en la Figura 1, la cual presenta una visión sintética de los seis estudios primarios identificados. En esta figura, los resultados han sido segmentados en dos áreas. La primera (lado izquierdo), consiste básicamente en dos gráficos (parte superior e inferior) de dispersión XY con burbujas en las intersecciones de las categorías tipo-año de publicación (lado izquierdo - parte superior) y tipo de publicación-contexto de usabilidad (lado izquierdo - parte inferior). Los tipos de publicación son conferencias y revistas. El tamaño de cada burbuja está determinado por el número de estudios primarios que se han clasificado como pertenecientes a cada categoría. La segunda área (lado derecho) de la Figura 1, presenta el número de estudios primarios por año de publicación (lado derecho - parte superior) y el número de métricas que han sido clasificadas en cada uno de los ocho contextos de usabilidad identificados

(lado derecho - parte inferior). Como se observa en la parte superior del lado derecho de la Figura 1, el interés en las métricas de usabilidad que pueden ser utilizadas en el desarrollo de apps ha aumentado a partir del 2016.

La segunda parte de esta sección presenta la síntesis del análisis de los estudios primarios, a partir del cual se respondió la pregunta de investigación.

RQ: ¿Qué métricas de accesibilidad y/o usabilidad, consideradas comúnmente en la construcción del software, podrían ser utilizadas para evaluar las apps desde la perspectiva de las PDV?

Como respuesta a la RQ fueron identificadas 73 métricas dentro de los estudios primarios. Las métricas se refieren o están relacionadas con la usabilidad en el desarrollo de aplicaciones móviles. Durante la extracción de las métricas, nos dimos cuenta que cada autor nombra de forma particular las métricas que estudia. Sin embargo, indistintamente de su denominación, las métricas se refieren a un contexto específico de usabilidad para dispositivos móviles. Por ello, para nuestra síntesis, hemos decidido agrupar las métricas identificadas en 8 contextos de usabilidad: Contenido visual, eficacia, navegabilidad, eficiencia, aprendizaje, satisfacción, comprensión y accesibilidad. Hemos codificado estos contextos como “CMA#” y a las métricas las hemos codificado como “CM#”.

No obstante, la agrupación permitió albergar 40 métricas en los diferentes contextos, como se muestra en la parte inferior del lado derecho de la Figura 1 (ver rectángulo con fondo gris y esquinas redondeadas). Por ejemplo, los contextos de Eficacia (CMA2) y Navegabilidad (CMA3) están compuestos por ocho métricas cada uno. Un total de 33 métricas no pudieron ser agrupadas en los contextos definidos, dado que su descripción no era clara. Las métricas agrupadas se encuentran en la **Tabla 4**. Los detalles de las métricas no agrupadas se pueden consultar en la sección “Metrics” de: <https://goo.gl/nmYxZw>.

En resumen, los resultados son interesantes. Por un lado, se verifica que todos los estudios primarios consideran por lo menos una de las métricas relacionadas con la eficacia (CMA2): Task Concordance (CM2), Usage characterization (CM5), Task used completely (CM25), No errors detected (CM26), Goal achieved correctly (CMA27), Effectiveness (CM62), Ease of Use (CM68), Effectiveness (CM71). Consideramos que estas métricas son importantes al tratarse de aplicaciones para dispositivos móviles. Por otro lado, las métricas más nombradas son: Visual coherence (CM1), graph properties (CM3), icons located correctly (CM13), image appeared correctly (CM14), text highlighted correctly (CM15), readability (CM61). Estas métricas podrían ser consideradas relevantes para apps que serán utilizadas por PDV, dada su relación muy cercana con temas de visualización.

5. Amenazas a la validez

El método de investigación utilizado en este trabajo corresponde a una versión reducida de una SLR, en la cual fueron consideradas algunas fases y etapas propuestas por Kitchenham & otros (Kitchenham, Brereton, O, & Linkman, 2009). A pesar de ser una revisión inicial de literatura, no fueron descuidados los criterios de rigurosidad para obtener estudios relevantes relacionados con la RQ. La **Tabla 5** muestra una evaluación

de la PLR, la cual fue realizada de acuerdo a los criterios sugeridos por Thompson otros (Thompson , Tiwari , Fu, Moe, & Buckley, 2012) , dado que son simples e independientes del dominio. Los resultados de la evaluación muestran que las decisiones tomadas en la PLR (basadas en su mayoría en validaciones cruzadas) fueron las adecuadas, lo que da muestras de la fiabilidad del método de investigación, y por ende de su validez.

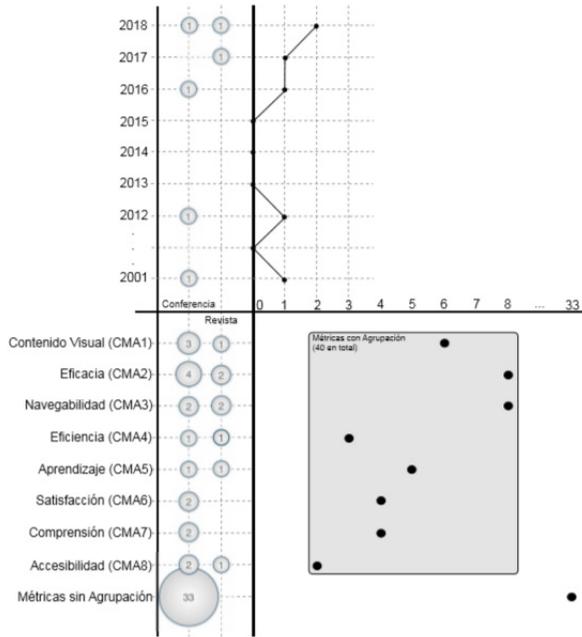


Figura 1 – Mapeo con la distribución de estudios primarios según los contextos de las métricas, incluyendo tipo, año de publicación y número de métricas por agrupación

Cod.	Agrupación	PS1	PS2	PS3	PS4	PS5	PS6
CMA1	Contenido visual	CM1: Visual coherence	CM3: Graph properties	CM13: Icons located correctly, CM14: Image appeared correctly, CM15: Text highlighted correctly		CM61: readability	
CMA2	Eficacia	CM2: Task Concordance	CM5: Usage characterization	CM25: Task used completely, CM26: No errors detected, CM27: Goal achieve correctly	CM68: Ease of Use	CM62: Effectiveness	CM71: Effectiveness
CMA3	Navegabilidad		CM6: Similarity	CM43: Input entering easily, CM44: Display self adjustment,	CM66: Navigation	CM65: Navigation	

		CM45: Task based on location accurately,		
		CM46: Less input error,		
		CM47: Surrounding interaction		
CMA4	Eficiencia	CM28: Task complete quickly,	CM64: Efficiency	
		CM29: Action done accurately		
CMA5	Aprendizaje	CM37: Skilladaptively,	CM67: Learnability	CM69: Learnability
		CM38: Learning to operate easily,		
		CM39: Task gain proficiency,		
CMA6	Satisfacción	CM48: Reach goal accurately,		CM73: Satisfaction
		CM49: Communication easily,		
		CM50: Icon button connect properly		
CMA7	Comprensión	CM58: Display load,		CM70: Understandability
		CM59: Operation clearly,		
		CM60: Operation completely,		
CMA8	Accesibilidad	CM30: Information Access easily	CM63: Accessibility	

Tabla 4 – Resumen de Métricas encontradas en los Estudios Primarios (PS)

Criterios de evaluación	Resultado Evaluación RQ
¿La búsqueda bibliográfica fue exhaustiva?	Si: Se realizaron varias iteraciones de selección aplicando filtros basados en los criterios de inclusión y exclusión.
¿Se utilizaron criterios apropiados en la selección de artículos para su inclusión?	Parcialmente: Los artículos seleccionados fueron apropiados, pero no especializados. Pudieron omitirse artículos que proponían métricas de usabilidad orientadas a las PDV.
¿Se incluyeron estudios que fueron suficientemente válidos para el tipo de pregunta formulada?	Si: Los artículos, tanto del grupo de control como los estudios primarios, abordaron específicamente el tema de métricas de usabilidad para el desarrollo de apps.
¿Los resultados fueron similares entre los estudios?	Si: fueron muy consistentes. Varias métricas se identificaron simultáneamente en dos o tres estudios primarios

Tabla 5 – Criterios para evaluar la PLR

6. Discusión, Conclusiones y Trabajos Futuros

Este trabajo describe una PLR realizada para responder a la pregunta de investigación: ¿Qué métricas de accesibilidad y/o usabilidad, consideradas comúnmente en la construcción del software, podrían ser utilizadas para evaluar las apps des-de la perspectiva de las PDV? Encontramos un total de 6 estudios primarios publicados en conferencias y revistas. Ninguno de estos estudios propone o plan-tea explícitamente métricas de usabilidad para el desarrollo de apps destinadas a las PDV.

Identificamos un total de 73 métricas de usabilidad reportadas en la literatura, de las cuales pudimos agrupar 40 en 8 contextos de usabilidad diferentes: contenido visual, eficacia, navegabilidad, eficiencia, aprendizaje, satisfacción, comprensión y accesibilidad. Los 3 primeros contextos agrupan la mayor cantidad de métricas de usabilidad.

Para las PDV es importante que las aplicaciones consideren principalmente 5 aspectos (Carvalho & Freire, 2017). En primer lugar, las apps deben proporcionar una retroalimentación adecuada para informar que una acción ha tenido un efecto. La retroalimentación está relacionada con los contextos de Comprensión y Contenido Visual. En segundo lugar, a partir de los elementos utilizados en la interfaz debe ser posible inferir correctamente la existencia de una funcionalidad. Es decir, los elementos de la interfaz de usuario utilizados (tales como, campos de entrada de texto, botones de opción) deben tener una asociación adecuada para que las PDV puedan inferir, por ejemplo, que están en una pantalla para el ingreso de datos (si este fuera el caso). Por tanto, este aspecto tiene relación con el contexto de Comprensión. En tercer lugar, la secuencia de interacción debe ser clara para que los usuarios comprendan los pasos que deben seguir para llevar a cabo una funcionalidad. Por ende, este aspecto tiene relación con los contextos de Comprensión y Navegabilidad. En cuarto lugar, debe ser posible realizar la identificación de los botones, de tal manera que cuando un botón específico sea tocado, el lector de pantalla no solamente lea “botón”, sino que también informe de la acción que realiza. Es decir, debe existir una asociación entre los botones y los elementos usados para etiquetarlos expresada en el código, de manera que las tecnologías de asistencia brinden una orientación adecuada. La identificación de los botones está relacionada con los contextos de Accesibilidad y Navegabilidad. Finalmente, las imágenes deben tener un texto alternativo cuando se utilice un lector de pantalla. De esta manera, es posible identificar si la imagen tiene asociada una acción y el usuario pueda, por ejemplo, iniciar una tarea. Este aspecto tiene relación con el contexto Eficacia. De lo anterior, podemos concluir que los contextos de usabilidad de mayor relevancia son: Contenido Visual, Eficacia, Navegabilidad, Comprensión y Accesibilidad. Los 3 primeros contextos coinciden con los que agrupan la mayor cantidad de métricas de usabilidad.

Nuestro estudio permitió establecer métricas de usabilidad que podrían ser adecuadas para el desarrollo de aplicaciones móviles dirigidas a las PDV. Sin embargo, es necesario realizar más esfuerzos de investigación en este sentido. Por tal razón, uno de los trabajos futuros consiste en evaluar algunas métricas (por ejemplo, navigation (CM65), information access easily (CM30)) en el desarrollo de una app para PDV. Otro de nuestros trabajos futuros es realizar una SLR completa y ampliar la búsqueda a otras bases digitales.

Agradecimientos: Esta investigación ha sido parcialmente financiada por el proyecto TIN 2014-60490-P (Ministerio de Economía y Competitividad, Gobierno de España), SENESCYT, ESPE y Armada del Ecuador.

Referencias

- Ahmad, N., Rextin, A., & Kulsoom, U. (2018). Perspectives on usability guidelines for smartphone applications: An empirical investigation and systematic literature review. *Information and Software Technology*, 130–149.
- Anju, S., & Sagar, K. (2017). A systematic review of software usability studies. *International Journal of Information Technology*, 1–24.
- Ba Matraf1, M. S., & Hussain, A. (2018). Modeling Measurement Metrics for E-Book App on Mobile Devices. *Journal of Telecommunication, Electronic and Computer Engineering*, 63–67.
- Baharuddin , R., Singh , D., & Razali , R. (2013). Usability dimensions for mobile applications. *Research Journal of Applied Sciences, Engineering and Technology*, 2225–2231.
- Baron, C. (2018). *The Statistics Portal*. Obtenido de The Statistics Portal: <https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/>
- Baumel, A., Faber, K., Mathur, N., Kane, J. M., & Muench, F. (2017). Enlight: A Comprehensive Quality and Therapeutic Potential Evaluation Tool for Mobile and Web-Based eHealth Interventions. *Journal of medical internet research*.
- Bolderston, A. (2008). Writing an Effective Literature Review. *Journal Medical Imaging and Radiation Scien*, 86–92.
- Campanella, P. (2012). Mobile Learning: New forms of education. *ICETA 2012 • 10th IEEE International Conference on Emerging eLearning Technologies and Applications*, 51–56.
- Carvajal, L., Moreno, A., & M. Segura, S. (2013). Usability through Software Design. *IEEE Transactions on Software Engineering*, 1582–1596.
- Carvalho, L. P., & Freire, A. P. (2017). Native or Web-Hybrid Apps?: An Analysis of the Adequacy for Accessibility of Android Interface Components Used with Screen Readers. *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, 38:1–38:10.
- Cruz Reis, R., Gomes, L., Dias-Neto, A., & de L. Fontão, A. (2015). Usability Evaluation Approaches for (Ubiquitous) Mobile Applications: A Systematic Mapping Study. *The Ninth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM)*.
- Dias Silva, F., Carvalho Nepomuceno, M. C., Silva, A. G., & Freire Pimenta, A. (2018). Accessibility and Usability Problems Encountered on Websites and Applications in Mobile Devices by Blind and Normal-vision Users. *Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC)*, págs. 2022–2029.

- Facanha Rocha, A., Cavalcante, M., & Campos, B. (2014). Touchscreen Mobile Phones Virtual Keyboarding for People with Visual Disabilities. *Human-Computer Interaction. Applications and Services: 16th International Conference, HCI International 2014, Heraklion*, 134–145.
- FortesPontin , R. M., Salgado, A. d., de Souza Santos, F., Agostini do Amaral, L., & Nogueira da Silva , E. A. (2017). Game Accessibility Evaluation Methods: A Literature Survey. *Springer International Publishing*, 182–192.
- Guerrero, H., & Vega, V. (2018). Usability analysis: Is our software inclusive. *Springer International Publishing AG*.
- Hanlu , Y., Malu, M., & Findla, U. (2014). Current and Future Mobile and Wearable Device Use by People with Visual Impairments. *Proceedings of the Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 3123–3132.
- Joorabchi, M. E., Mesbah, A., & Kruchten, P. (2013). Real Challenges in Mobile App Development. *2013 ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, 15–24.
- Kane, S. K. (2009). Freedom to Roam: A Study of Mobile Device Adoption and Accessibility for People with Visual and Motor Disabilities. *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 115–122.
- Kitchenham, B., Brereton, O , P., & Linkman, S. (2009). Systematic literature reviews in software engineering--a systematic literature review. *Information and software technology*, 7–15.
- Kronbauer H, A., & Santos S., C. A. (2012). Smartphone Applications Usability Evaluation: A Hybrid Model and Its Implementation. *Human-Centered Software Engineering*, 146–163.
- Kulpa Costa, C., & Amaral Gon, F. (s.f.). Avaliação da interação entre usuários de baixa visão e as interfaces gráficas digitais de um tablet: foco na usabilidade. *Blucher Design Proceedings. Blucher*, 2015.
- Kumanan , W., Atkinson , K., Westeinde , J., Bell, C., Marty , K., Dean , F., . . . Bett, J. (2016). An evaluation of the feasibility and usability of a proof of concept mobile app for adverse event reporting post influenza vaccination. *Human Vaccines \& Immunotherapeutics*, 1738–1748.
- Kumar, B. A., & Mohite, P. (2018). Usability of mobile learning applications: a systematic literature review. *Journal of Computers in Education*, 1–17.
- Lettner, F., & Holzmann, C. (2012). Automated and Unsupervised User Interaction Logging As Basis for Usability Evaluation of Mobile Applications. *Proceedings of the 10th International Conference on Advances in Mobile Computing \& Multimedia*, 118–127.
- Molano I., J. R., Yara S. , E., & García K. , L. J. (2015). Model for measuring usability of survey mobile apps, by analysis of usability evaluation methods and attributes. *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6.

- Olmedo-Moreno , E. M., & López-Delgado , A. (2015). Apps. accessibility and usability by people with visual disabilities. *12th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA)* , 267–270.
- P. Di , G., M., R., M. , S., G. , T., G. , V., T. , G., . . . G. , W. (2012). User centered scenario based approach for developing mobile interfaces for Social Life Networks. *First International Workshop UsARE*, 18–24.
- Sánchez Sierra, J., & Togores, J. (2012). Designing mobile apps for visually impaired and blind users. *The Fifth International Conference on Advances in Computer-Human Interactions*, 47–52.
- Sauro, J., & Lewis R., J. (2009). Correlations Among Prototypical Usability Metrics: Evidence for the Construct of Usability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1609–1618.
- Seffah , A., Kececi, N., & Donyaee, M. (2001). QUIM: A Framework for Quantifying Usability Metrics in Software Quality Models. *Proceedings Second Asia-Pacific Conference on Quality Software*, 311–318.
- Thompson , M., Tiwari , A., Fu, R., Moe, E., & Buckley, D. I. (2012). A Framework to Facilitate the Use of Systematic Reviews and Meta-analyses in the Design of Primary Research Studies. *Agency for Healthcare Research and Quality - Department of Health and Human Services*.
- TIC, O. A. (27 de Agosto de 2013). discapnet. *Fundación Once*. Obtenido de disc@pnet: http://www.discapnet.es/sites/default/files/areas-tematicas/tecnologia/informe_sintetico_observatorio_aplicaciones_moviles_27-08-2013.pdf
- Velásquez, J. (2015). Una Guía Corta para Escribir Revisiones Sistemáticas de Literatura. *Research Gate*.
- World Health Organization. (2017). Obtenido de Blindness and visual impairment: <http://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Zali , Z., & Ahmad Fadzlah , A. F. (2016). An initial theoretical usability evaluation model for assessing defence mobile e-based application system . *International Conference on Information and Communication Technology (ICICTM)*.
- Zhanga, H., Ali Babarb, M., & Tellb, P. (2010). Identifying Relevant Studies in Software Engineering. *National ICT Australia, University of New South Wales, Australia*.

© 2019. This work is published under <https://creativecommons.org/licenses/by-nc-nd/4.0/>(the “License”). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License.