

Estudio de métodos de evaluación de usabilidad colaborativos en el área web transaccional

Study of collaborative usability evaluation methods in transactional web area

Andrés Solano, César A. Collazos
Departamento de Sistemas, FIET
Universidad del Cauca
Popayán, Colombia
{afsolano, ccollazo}@unicauca.edu.co

Cristian Rusu
Escuela de Ingeniería Informática
Pontificia Universidad Católica de Valparaíso
Valparaíso, Chile
cristian.rusu@ucv.cl

Resumen—La usabilidad es una característica de calidad fundamental para el éxito de un sistema interactivo. La usabilidad incluye una serie de métodos de evaluación con el objetivo de obtener sistemas eficientes, fáciles de usar y de aprender. Los métodos de evaluación de usabilidad son bastante diversos y su realización depende de variables tales como: costos, disponibilidad de tiempo, entre otras. En ese sentido, este artículo presenta el análisis de los resultados obtenidos a partir de la ejecución de un conjunto de métodos de evaluación de usabilidad diseñados de forma colaborativa mediante el uso de una metodología que provee la *Ingeniería de Colaboración*. Dicho análisis está basado en una serie de métricas relacionadas con la detección de problemas y tiempo. Esto con el fin de identificar los métodos de evaluación que permiten obtener resultados más apropiados en el área de web transaccional.

Abstract—Usability is a key quality feature to the success of interactive system. Usability includes a set of metrics and evaluation methods in order to obtain efficient systems, easy to use and easy to learn. Usability evaluation methods are quite diverse; its realization depends of variables like costs, time availability, among others. In that sense, this paper presents the analysis of the obtained results from the development of a set of usability evaluation methods designed collaboratively using a methodology provided by the *Collaboration Engineering* approach. The analysis is based on a set of metrics related to usability problems detection and time. This in order to identify the usability evaluation methods that obtain more appropriate results in transactional web area.

Palabras Claves—*Interacción Hombre-Máquina; usabilidad; métodos de evaluación de usabilidad; métricas; web transaccional; Ingeniería de Colaboración.*

Keywords—*Human-Computer Interaction; usability; usability evaluation methods; metrics; transactional web; Collaboration Engineering.*

I. INTRODUCCIÓN

Sin lugar a dudas, actualmente la tendencia cada vez más común es la de trabajar colaborativamente entre personas para alcanzar un objetivo común. El trabajo se organiza en equipos

y cada integrante interactúa con el resto del grupo para obtener una mejor productividad [1]. Al integrar aspectos de trabajo colaborativo a un proceso determinado, el objetivo no es sólo la mejora de la comunicación, sino también lograr mayor participación, compromiso, entre los integrantes de un grupo que trabajan en torno a una actividad común, lo que conlleva a una mejor calidad del producto elaborado [2].

Centrándonos en el contexto del diseño y evaluación de interfaces de usuario, el proceso de evaluación de usabilidad no es ajeno a esta tendencia de trabajar colaborativamente. Desde siempre, la disciplina *Interacción Humano-Computador* reconoce la necesidad de equipos multidisciplinares que permitan realizar una evaluación de forma más adecuada. Así, con el objetivo de mejorar el proceso tradicionalmente definido, la *Metodología para el Desarrollo de Procesos Colaborativos* (MDPC) [3] ha sido utilizada para obtener la especificación colaborativa de un conjunto de Métodos de Evaluación de Usabilidad (MEU). En dicha especificación son definidos procesos colaborativos (en los que participan varias personas de diversas áreas de conocimiento, las cuales pueden estar distribuidas geográficamente), los roles de los miembros del grupo, el proceso de comunicación, entregables a generar, entre otra información [4].

Por otro lado, con el rápido crecimiento de la web, el número de aplicaciones E-Commerce está continuamente en aumento. De esta manera, la usabilidad juega un rol importante pues las aplicaciones web transaccionales deben permitir a los usuarios lograr sus objetivos con efectividad, eficiencia y satisfacción, siempre teniendo en cuenta que las aplicaciones deben ser comprensibles y fáciles de usar [5]. Las aplicaciones web transaccionales son sistemas software particularmente complejos, en donde existe una interacción constante con el usuario y pueden ser repositorios de gran cantidad de información, además de ofrecer servicios que pueden ser accedidos desde cualquier lugar y desde variados dispositivos [6]. Ejemplos de sistemas relacionados a la web transaccional son: sistemas bancarios, compras en línea, sistemas de reservas, entre otras, los cuales permiten al usuario realizar transacciones (no necesariamente financieras) [6].

Ahora bien, los MEU que permiten medir la aplicación de éste atributo en cierto sistema y bajo ciertos factores, son bastante diversos [7]. Su realización depende de variables tales como: costos, disponibilidad de tiempo, entre otras. Teniendo en cuenta lo anterior, este trabajo está enfocado en el estudio de un conjunto de MEU (especificados de forma colaborativa mediante la MDPC) sobre una aplicación web transaccional, analizando los resultados obtenidos a partir de una serie de métricas relacionadas con la detección de problemas y tiempo. Esto con el fin de identificar los MEU más apropiados para ejecutar en el área web transaccional.

La sección 2 presenta los MEU y la aplicación web transaccional objeto de estudio. La descripción de las métricas es presentada en la sección 3. La sección 4 presenta los resultados obtenidos para las métricas definidas a partir de la ejecución de los MEU, y la sección 5, presenta el análisis de dichos resultados. Finalmente, la sección 6 presenta algunas conclusiones y trabajos futuros.

II. MÉTODOS DE EVALUACIÓN DE USABILIDAD Y APLICACIÓN WEB TRANSACCIONAL

A. Selección de los MEU y aplicación web objeto de estudio

Dado que existe un buen número de métodos para evaluar la usabilidad de sistemas interactivos, se hace necesario seleccionar un conjunto reducido de métodos objeto de estudio. Para ello, se ha tomado como referente el trabajo doctoral [8], en el cual ha sido realizada una *valoración total de utilidad* de los MEU (asignando uno de los siguientes valores: *muy útil, útil y poco útil*) en base a criterios como: aplicabilidad, necesidad de formación, representatividad, aportación vs esfuerzo, entre otros. Así, fueron considerados los métodos de evaluación valorados como “muy útiles” y “útiles”. Adicionalmente, en el proceso de selección de los MEU objeto de estudio fueron consideradas las debilidades y fortalezas de los diferentes métodos, así como también las experiencias y recomendaciones consignadas en trabajos de investigación relacionados. De esta manera, los MEU (de inspección y prueba) seleccionados son: **(1) evaluación heurística, (2) recorrido cognitivo, (3) experimentos formales, (4) cuestionarios, (5) interacción constructiva, (6) entrevistas y (7) método del conductor.**

Para la elección de la aplicación web transaccional fueron considerados los siguientes criterios: (i) *Representatividad*: el sistema debe ser representativo del área de aplicación a la que pertenece, en este caso web transaccional. (ii) *Disponibilidad*: el sistema debe estar disponible de forma gratuita. Además, el sistema debe ser de fácil acceso. (iii) *Tareas representativas en el sistema*: el sistema interactivo debe contar con una cantidad apropiada de funcionalidades y un buen nivel de navegabilidad, así será posible realizar tareas representativas.

Considerando los criterios antes mencionados, el sitio web transaccional a evaluar es **Booking.com** (disponible en la URL: www.booking.com), el cual es un sistema final que ofrece el servicio de reservas hoteleras y funciona a nivel internacional. Este sistema está disponible de forma gratuita, es de fácil acceso y presenta una cantidad apropiada de funcionalidades con buen nivel de navegabilidad. Booking.com

fue seleccionado porque una reserva hotelera es un ejemplo representativo de transacción, por lo cual los resultados obtenidos en las evaluaciones podrían generalizarse a otros sistemas que pertenecen al área web transaccional.

B. Ejecución de los MEU sobre la aplicación web

Los métodos de inspección fueron realizados en el lugar de trabajo de los evaluadores puesto que estos se encontraban distribuidos geográficamente. La evaluación heurística fue realizada por un conjunto de 5 evaluadores que inspeccionaron el diseño de la interfaz de la aplicación con base en una serie de principios específicos para web transaccional [5]. Este método de inspección permitió detectar 36 problemas de usabilidad, siendo el principio *Control y libertad de usuario* el que tuvo mayor cantidad de problemas asociados. En general, el nivel de criticidad de los problemas es bajo, una poca cantidad de problemas (6 de 36) fueron calificados, en promedio, con notas mayores a 6 (en una escala de 0 a 8). El recorrido cognitivo fue realizado por un conjunto de 4 evaluadores que realizaron una serie de tareas con el fin de comprobar si las interfaces son adecuadas para los usuarios.

Los métodos de prueba (con usuarios representativos que se adecuan al perfil de usuario definido) fueron realizados en las instalaciones de la Universidad del Cauca (Colombia). En los experimentos formales el listado de tareas relacionado a las funcionalidades objeto de estudio fue elaborado con base en los problemas críticos identificados en la evaluación heurística. Por otro lado, los cuestionarios fueron realizados luego de llevar a cabo los experimentos formales. Con la ejecución de los cuestionarios no fueron identificados problemas de usabilidad, sin embargo, se obtuvo una serie de cálculos estadísticos con base en las respuestas de los usuarios, relacionados a la satisfacción subjetiva de los mismos. En la interacción constructiva las parejas de usuarios realizaron una exploración “libre” del sistema al mismo tiempo que expresaban sus impresiones a viva voz. Debido a la naturaleza del método, fueron realizadas 5 interacciones. Las entrevistas fueron realizadas luego de las interacciones constructivas, esto con el fin de obtener información acerca de la percepción de los usuarios acerca de distintos aspectos de la aplicación. Por último fue realizado el método del conductor en el cual los usuarios fueron guiados (con escenario de uso preestablecido) en la realización de una serie de tareas. En resumen, la Tabla 1 presenta la cantidad de problemas identificados por los MEU.

TABLA 1. IDENTIFICACIÓN DE PROBLEMAS POR LOS MEU.

MEU	# Usuarios	# Problemas	# Problemas críticos confirmados	# Problemas no detectados en la evaluación heurística
Evaluación heurística	No aplica	36	6	No aplica
Recorrido cognitivo	No aplica	21	6	7
Experimentos formales	10	24	6	11
Cuestionarios	10	0	0	0
Interacción constructiva	10	29	6	7
Entrevistas	10	14	5	3
Método del conductor	8	25	6	9

Una vez ejecutados los MEU sobre la aplicación web transaccional, fueron analizados los resultados obtenidos. Para el análisis de resultados fue definido un conjunto de métricas las cuales son descritas en la siguiente sección.

III. DESCRIPCIÓN DE LAS MÉTRICAS

A. Identificación de las métricas

Para el análisis de los resultados obtenidos en la ejecución de los MEU es necesario definir un conjunto de métricas que permitan medir de forma objetiva dichos resultados obtenidos. Para ello, luego de un proceso de observación y revisión de la literatura, se obtuvo una serie de métricas las cuales fueron agrupadas según las siguientes características: *detección de problemas de usabilidad*, *recurso humano*, *equipamiento*, *tiempo* y *tareas*. Luego de definir el conjunto preliminar de métricas, fue elaborada una encuesta con el fin de identificar, según la experiencia y conocimiento de un grupo de expertos, las métricas más relevantes para llevar a cabo el análisis de resultados. La encuesta fue elaborada utilizando el sistema SUS (*System Usability Scale*) [9], de tal manera que cada pregunta tiene 5 opciones de respuesta. Así, fue realizado un consenso entre 11 participantes con experiencia en la evaluación de usabilidad de sistemas interactivos.

B. Selección de las métricas

Una vez recopilados y procesados los resultados de las encuestas (incluyendo promedios y desviación estándar), fueron seleccionadas las métricas más relevantes según sus altos promedios. Dichas métricas son aquellas calificadas como “importantes” y “muy importantes” con base en la experiencia de los participantes que diligenciaron las encuestas. Las métricas seleccionadas pertenecen a las características: *detección de problemas de usabilidad*, *recurso humano* y *tiempo*. Sin embargo, al hacer el análisis entre los MEU deben tenerse en cuenta las métricas generadas por los propios métodos, por lo cual las métricas de la característica *recurso humano* no son consideradas como criterios para discriminar entre los MEU objeto de estudio. La razón es que dichas métricas no están relacionadas con el método en sí, sino con una sesión de prueba en la que este es usado, lo cual es diferente. Por ejemplo, el número de personas que intervienen

en la ejecución de un método (métrica *Cantidad de implicados*) no debería ser un criterio para comparar entre varios MEU pues se estaría atribuyendo a una métrica un valor que no es generado por el propio método (los requisitos del método o las exigencias de trabajo del mismo). Lo mismo sucede con la métrica *Experiencia (en años) de expertos/evaluadores*, puesto que no revela nada sobre el método. La Tabla 2 presenta la descripción de las métricas seleccionadas.

Las métricas seleccionadas corresponden a medidas base (o directas), esto indica que no dependen de ninguna otra medida [9]. Las métricas que pertenecen a la característica *detección de problemas de usabilidad* están asociadas a un tipo de *escala absoluta* [10] ya que sólo hay una forma posible de medir: contando; mientras que las métricas de la característica *tiempo* están asociadas a un tipo de *escala de ratio* [10], la cual tiene un punto fijo de referencia: el cero (ningún valor puede ser menor a cero). Ahora bien, una vez realizada la medición, los valores de las métricas no están entre 0 y 1 (exceden a 1), por lo cual debe ser utilizada una tabla de “normalización” para llevarlos a escala de valores entre 0 y 1. Luego de normalizar los valores, las métricas generan un número real que está entre 0 y 1. Así, las métricas proporcionan evidencias positivas si los valores son cercanos a 1. Cabe mencionar que para métricas cuyos valores “buenos” son los cercanos a cero, habría necesidad de efectuar un cálculo como este: $V_c = 1 - V$. De modo que cuando el valor (V) de la métrica se acerque más a cero, el valor complementario (V_c) será más cercano a 1 de modo que todas las métricas puedan llevarse a valores con sentido positivo (o creciente).

Respecto a las métricas asociadas a la característica *tiempo*, no se ha establecido un tiempo base para realizar las etapas de planeación, ejecución y análisis de resultados. La razón es que el tiempo puede variar según el número de evaluadores y usuarios que participen en el proceso de evaluación. Por otro lado, la velocidad de ejecución de las etapas de un MEU (planeación, ejecución y análisis de resultados) no es un valor por sí misma si no se asegura una mínima calidad en los resultados que se deriven de ellas. Por lo cual el *evaluador supervisor* (rol definido en la especificación colaborativa de los MEU) es el encargado de asegurar cierto grado de calidad en los resultados/entregables obtenidos.

TABLA 2. DESCRIPCIÓN DE LAS MÉTRICAS A CONSIDERAR EN EL ANÁLISIS DE RESULTADOS DE LOS MEU.

Métrica	Descripción e interpretación
Característica: Detección de problemas de usabilidad	
Cantidad total de problemas identificados (CTP)	Esta métrica corresponde a la cantidad total de problemas de usabilidad identificados en el sistema evaluado. Cuantos más problemas de usabilidad sean detectados, el valor de la métrica se acerca más a 1.
Cantidad de problemas críticos (CPC)	Esta métrica corresponde a la cantidad de problemas críticos identificados en el sistema evaluado. Cuantos más problemas críticos sean detectados, el valor de la métrica se acerca más a 1.
Cantidad de problemas frecuentes (CPF)	Esta métrica corresponde a la cantidad de problemas frecuentes identificados en el sistema evaluado. Cuantos más problemas frecuentes sean detectados, el valor de la métrica se acerca más a 1.
Característica: Tiempo	
Tiempo empleado para completar la planeación (TEP)	Esta métrica corresponde al tiempo empleado para realizar las actividades que conforman la <i>etapa de planeación</i> . Cuanto menos tiempo sea empleado para realizar las actividades de la <i>etapa de planeación</i> , mejor. Habría que definir: $TEP_c = 1 - TEP$, para definir relación: menor (tiempo) – mayor (valor).
Tiempo empleado para completar la ejecución (TEE)	Esta métrica corresponde al tiempo empleado para realizar las actividades que conforman la <i>etapa de ejecución</i> . Cuanto menos tiempo sea empleado para realizar las actividades de la <i>etapa de ejecución</i> , mejor. Habría que definir: $TEE_c = 1 - TEE$, para definir relación: menor (tiempo) – mayor (valor).
Tiempo empleado para completar el análisis de resultados (TEA)	Esta métrica corresponde al tiempo empleado para completar el análisis de resultados. Cuanto menos tiempo sea empleado para analizar los resultados, mejor. Habría que definir: $TEA_c = 1 - TEA$, para definir relación: menor (tiempo) – mayor (valor).

C. Normalización de las métricas

Para las métricas que pertenecen a la característica *detección de problemas de usabilidad y tiempo*, el número de intervalos fue establecido siguiendo algunas pautas definidas en [11]. Además, las *tablas de normalización* toman como valores de referencia (valores máximos) la cantidad de problemas detectados (críticos y frecuentes) en la evaluación heurística. Este método de evaluación es la base de comparación para el análisis de los métodos de prueba, debido a que es uno de los más eficientes y más utilizados [2][7].

Cabe resaltar que la medida de las métricas asociadas a la característica *tiempo* es una aproximación, puesto que el tiempo para completar las actividades que conforman los MEU puede variar según el número de evaluadores y usuarios que participen en el proceso de evaluación, número de funcionalidades a evaluar, número de tareas diseñadas para los usuarios, entre otros factores. El tiempo en las actividades que son realizadas varias veces (como por ejemplo: la presentación de una prueba, la realización de tareas en el sistema por parte de los usuarios, la realización de preguntas para obtener información adicional, entre otras) corresponde al tiempo promedio que demora el *evaluador supervisor* o *usuario(s)* realizando la actividad en cuestión. De igual manera, el tiempo de las actividades colaborativas corresponde al tiempo promedio que cada evaluador invierte en la ejecución de los pasos propuestos en el(los) *thinklet(s)* [3] respectivo(s).

IV. RESULTADOS

Esta sección presenta las medidas de las métricas según los resultados obtenidos en la ejecución de los MEU. Respecto a los métodos de medición, cabe mencionar que no se emplearon herramientas software que los automatizaran. Para conocer la medida de las métricas fue realizado un conteo manual, a partir de un seguimiento observacional durante la ejecución de los MEU (de las grabaciones, documentos guía, entre otros).

A. Detección de problemas de usabilidad

Los valores de las métricas CTP, CPC y CPF no están entre 0 y 1, por lo cual son utilizadas *tablas de normalización*. Para el caso de la métrica CTP, la Tabla 3 presenta la tabla de normalización propuesta para dicha métrica la cual toma como valor de referencia la cantidad de problemas detectados en la evaluación heurística. De igual forma, las tablas de normalización para las métricas CPC y CPF (no presentadas en el presente artículo por restricciones de extensión) utilizan como valores de referencia la cantidad de problemas críticos y frecuentes detectados en la evaluación heurística.

Cantidad total de problemas identificados (CTP)

La Tabla 3 presenta la tabla de normalización propuesta para la métrica CTP.

TABLA 3. TABLA DE NORMALIZACIÓN PARA LA MÉTRICA CTP.

Clasificación del valor obtenido de la métrica	Calificación a asignar para "normalizar"
Valor Métrica > 36	0,95
30 < Valor Métrica <= 36	0,80
24 < Valor Métrica <= 30	0,65
18 < Valor Métrica <= 24	0,50

Clasificación del valor obtenido de la métrica	Calificación a asignar para "normalizar"
12 < Valor Métrica <= 18	0,35
6 < Valor Métrica <= 12	0,20
Valor Métrica <= 6	0,05

Con base en la Tabla 3, la Tabla 4 presenta las medidas obtenidas y normalizadas de la métrica CTP.

TABLA 4. MEDIDAS DE LA MÉTRICA CTP.

Método de evaluación	CTP	CTP normalizada
Evaluación heurística	36	0,80
Recorrido cognitivo	21	0,50
Experimentos formales	24	0,50
Cuestionarios	0	0,05
Interacción constructiva	29	0,65
Entrevistas	14	0,35
Método del conductor	25	0,65

Cantidad de problemas críticos y frecuentes (CPC y CPF)

La Tabla 5 presenta las medidas obtenidas y normalizadas de las métricas CPC y CPF.

TABLA 5. MEDIDAS DE LAS MÉTRICAS CPC Y CPF.

Método de evaluación	CPC	CPC normalizada	CPF	CPF normalizada
Evaluación heurística	6	No aplica	12	No aplica
Recorrido cognitivo	6	1	10	0,84
Experimentos formales	6	1	8	0,67
Cuestionarios	0	0	0	0
Interacción constructiva	6	1	11	0,84
Entrevistas	5	0,75	8	0,67
Método del conductor	6	1	9	0,67

B. Tiempo

Los valores de las métricas TEP, TEE y TEA no están entre 0 y 1, por lo cual es utilizada una *tabla de normalización* para llevarlos a escala de valores entre 0 y 1. Además, debe calcularse el complemento para definir la relación: menor (tiempo) – mayor (valor). La Tabla 6 presenta la *tabla de normalización* propuesta (en minutos) para las métricas asociadas a la característica *tiempo*.

Es importante mencionar que las actividades colaborativas relacionadas a la etapa de planeación, ejecución y análisis de resultados, fueron realizadas de forma virtual (utilizando documentos compartidos como herramienta ofimática colaborativa) debido a que los evaluadores se encontraban distribuidos geográficamente. Por lo cual, las contribuciones por parte de los evaluadores fueron recolectadas en un periodo de tiempo considerable (4 días en promedio).

TABLA 6. TABLA DE NORMALIZACIÓN PARA LAS MÉTRICAS CORRESPONDIENTES A LA CARACTERÍSTICA TIEMPO.

Clasificación del valor obtenido de la métrica (minutos)	Calificación a asignar para "normalizar"
Tiempo Actividades > 450	0,99
420 < Tiempo Actividades <= 450	0,92
390 < Tiempo Actividades <= 420	0,85
360 < Tiempo Actividades <= 390	0,78
330 < Tiempo Actividades <= 360	0,71
300 < Tiempo Actividades <= 330	0,64
270 < Tiempo Actividades <= 300	0,57
240 < Tiempo Actividades <= 270	0,50

Clasificación del valor obtenido de la métrica (minutos)	Calificación a asignar para "normalizar"
210 < Tiempo Actividades <= 240	0,43
180 < Tiempo Actividades <= 210	0,36
150 < Tiempo Actividades <= 180	0,29
120 < Tiempo Actividades <= 150	0,22
90 < Tiempo Actividades <= 120	0,15
30 < Tiempo Actividades <= 90	0,08
0 < Tiempo Actividades <= 30	0,01

Tiempo empleado para completar la planeación (TEP)

La Tabla 7 presenta las medidas de la métrica TEP. Las medidas de las métricas TEE y TEA no se presentan debido a restricciones de extensión del documento.

TABLA 7. MEDIDAS DE LA MÉTRICA TEP.

Método de evaluación	TEP (minutos)	Valor normalizado	Complemento de TEP
Evaluación heurística	358	0,71	0,29
Recorrido cognitivo	292	0,57	0,43
Experimentos formales	455	0,99	0,01
Cuestionarios	122	0,22	0,78
Interacción constructiva	142	0,22	0,78
Entrevistas	45	0,08	0,92
Método del conductor	209	0,36	0,64

V. ANÁLISIS DE RESULTADOS

Con base en los resultados obtenidos, y considerando los factores bajo los cuales fueron ejecutados los MEU, se tiene el siguiente análisis de resultados.

A. Detección de problemas de usabilidad

El método con la calificación más alta respecto a la CTP es la evaluación heurística. Otros métodos que también obtuvieron calificaciones altas son la interacción constructiva y el conductor, puesto que permitieron identificar una cantidad significativa de problemas. El método con la calificación más baja se refiere a los cuestionarios pues no fueron identificados problemas de usabilidad mediante su ejecución, sin embargo, estos proporcionaron una serie de cálculos estadísticos relacionados a la satisfacción subjetiva de los usuarios. En cuanto a la CTP, la evaluación heurística e interacción constructiva permitieron identificar la mayor cantidad de problemas de usabilidad, puesto que en estos métodos el sistema es evaluado de forma global. Esto es, los evaluadores (en la evaluación heurística) y las parejas de usuarios (en la interacción constructiva) exploran todas las funcionalidades que ofrece el sistema, con lo cual el número de problemas identificados se incrementa en buena medida. Por otro lado, se tienen los métodos: recorrido cognitivo, experimentos formales y conductor, los cuales utilizan una lista de tareas para realizar en el sistema, por tal razón, el número de problemas detectados se minimiza (en comparación a los métodos que hacen una evaluación global) pues estos corresponden a las funcionalidades evaluadas. Cabe destacar que el método del conductor permitió identificar la mayor cantidad de problemas entre aquellos que utilizan una lista de tareas específicas.

Respecto a la métrica CPC, la mayoría de métodos confirmaron los problemas críticos identificados en la evaluación heurística. Los métodos que obtuvieron las calificaciones más bajas son las entrevistas y cuestionarios, ya

que estos se enfocaron en obtener información acerca de las impresiones de los usuarios y la satisfacción subjetiva respecto al uso del sistema. Sin embargo, las entrevistas y cuestionarios funcionaron como buen complemento de otros métodos de prueba para capturar información adicional por parte de los usuarios, a pesar de que no permitieron identificar un número significativo de problemas de usabilidad.

Los métodos con las calificaciones más altas respecto a la CPF son la interacción constructiva y recorrido cognitivo. Los demás métodos (excepto los cuestionarios) también obtuvieron calificaciones altas ya que permitieron detectar un buen número de los problemas frecuentes identificados en la evaluación heurística. Al igual que en las métricas CTP y CPC, el método con la calificación más baja se refiere a los cuestionarios.

Con base en lo anterior, la interacción constructiva y conductor, a pesar de que presentan un nivel considerable de subjetividad, presentan resultados positivos por encima de los experimentos formales, lo cual podría indicar que se obtienen resultados (realimentación) apropiados con métodos de evaluación que promueven una interacción directa con los usuarios. Dichos métodos posiblemente son llevados a cabo en un contexto que no es el más real, ya que en una aplicación web transaccional generalmente no interactúan dos usuarios a la vez. Sin embargo, son métodos que funcionan muy bien y permiten recolectar buena cantidad de información acerca de las percepciones subjetivas de los usuarios, ya que estos se sienten con más confianza de expresar sus impresiones.

B. Tiempo

El método con la calificación más alta respecto al TEP se refiere a las entrevistas, las cuales no necesitaron una extensa preparación y fue posible prepararlas en menos de una hora. Otros métodos que obtuvieron una calificación media son los cuestionarios y la interacción constructiva, pues su preparación requirió en promedio de 2 a 3 horas. El método con la calificación más baja corresponde a los experimentos formales, puesto que la planeación de este método incluye una mayor cantidad de actividades (relacionadas al diseño de tareas) en comparación a los demás MEU objeto de estudio.

Los métodos con las calificaciones más altas respecto al TEE son las entrevistas y cuestionarios. Por lo cual, el tiempo promedio invertido por los evaluadores y usuarios realizando las actividades que conforman la *etapa de ejecución* de dichos métodos, es reducido. Otros métodos con una calificación media/alta son: conductor, experimentos formales y la interacción constructiva. El método más demorado fue la evaluación heurística debido a que requirió de un período de tiempo considerable para que los evaluadores inspeccionen el sistema tranquila y detalladamente.

Respecto a la métrica TEA, los métodos que obtuvieron las calificaciones más altas son las entrevistas y el recorrido cognitivo. Esto sugiere que el tiempo promedio invertido por los evaluadores en las actividades de análisis es relativamente bajo, pues la información recolectada en las pruebas (grabaciones, documentos guía y demás anotaciones) es distribuida entre el grupo de evaluadores. En el caso de las entrevistas, cabe mencionar que el poco tiempo demorado en el

análisis de resultados no necesariamente estuvo relacionado con las características del método, sino a la disponibilidad limitada de los evaluadores para participar en las actividades sugeridas. Por otro lado, el método más demorado en el análisis corresponde a los experimentos formales, ya que cada evaluador invirtió un tiempo considerable analizando la información (2 o 3 grabaciones) distribuida por el *evaluador supervisor*. Cabe mencionar que el poco tiempo demorado en el análisis de resultados no necesariamente se debe a características de los MEU, sino al poco tiempo disponible de los evaluadores para participar en las actividades. Hubo casos en los cuales algunos evaluadores no participaron de todos los pasos propuestos en las actividades.

VI. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo fue ejecutado un conjunto de MEU sobre una aplicación que pertenece al área web transaccional, presentando el análisis de los resultados obtenidos con base en una serie de métricas. Es importante recalcar que no existe un “mejor método”, todos tienen fortalezas y debilidades y están enfocados a evaluar aspectos específicos de usabilidad, por tal razón, según [8], combinarlos es el procedimiento más apropiado. De esta manera, se propone, en primer lugar, una combinación de métodos para una **evaluación global** que permita detectar un alto número de problemas de usabilidad, dicha combinación está conformada por los métodos: **evaluación heurística + interacción constructiva + entrevistas**. En segundo lugar, se propone una combinación de métodos para la **evaluación específica** de un conjunto de funcionalidades o áreas de un sistema interactivo, dicha combinación está conformada por los métodos: **evaluación heurística + método del conductor + cuestionarios**.

Considerando el estado actual de la literatura, son escasos los trabajos relacionados que incluyen aspectos de trabajo colaborativo a los procesos de evaluación de usabilidad. En [6] se han propuesto una serie de combinaciones de métodos para evaluar la usabilidad de aplicaciones web transaccionales, por lo que este trabajo contribuye a dicha propuesta pues contempla MEU adicionales y ha generado la especificación colaborativa de los MEU objeto de estudio.

En las métricas relacionadas a la característica *tiempo* es discutible si un valor “pequeño” constituye un hecho positivo pues podría dudarse de la calidad de los entregables generados en cada actividad. De igual forma, la velocidad de ejecución de las etapas de un método de evaluación (planeación, ejecución y análisis de resultados) no es un valor por sí misma si no se asegura una mínima calidad en los resultados que se deriven de ellas. Por tal razón, el *evaluador supervisor* (rol definido en la especificación colaborativa de los MEU) es el encargado de asegurar cierto grado de calidad en los resultados obtenidos.

Teniendo en cuenta la métrica: *Experiencia (en años) de expertos/evaluadores*, en la evaluación heurística y recorrido cognitivo participaron evaluadores con un nivel de experiencia medio/alto (más de 5 años en promedio) en evaluaciones de usabilidad de sistemas interactivos. En los demás métodos de evaluación participaron evaluadores con un nivel de experiencia medio/bajo (menor o igual a 4 años en promedio).

Considerando lo anterior, resulta complicado asegurar que sean los MEU los causantes de un mayor rendimiento a la hora de detectar problemas de usabilidad porque en algunos métodos han participado más o menos evaluadores con más experiencia que otros. Sin embargo, en la vida real no es tarea fácil conseguir evaluadores con experiencia alta en la evaluación de sistemas interactivos, que conozcan las características del dominio del sistema y estén disponibles. Claramente contar con evaluadores experimentados permitirá obtener resultados más apropiados e incrementará la confiabilidad de los mismos.

Como actividad futura, conviene llevar a cabo otros casos de estudio en los que participen evaluadores con el mismo nivel de experiencia con el fin de aislar aquellos factores que pueden tener influencia en el estudio.

AGRADECIMIENTOS

Este trabajo ha sido apoyado por el grupo de Investigación y Desarrollo en Ingeniería de Software (IDIS) de la Universidad del Cauca (Colombia), y por el grupo de investigación en Interacción Humano-Computador (useCV) de la Pontificia Universidad Católica de Valparaíso (Chile). Ha sido parcialmente financiado por el Programa Nacional para Estudios de Doctorado en Colombia Año 2011, de Colciencias.

REFERENCIAS

- [1] C. A. Ellis, S. J. Gibbs, G. Rein, "Groupware: some issues and experiences," *Communications of the ACM*, vol. 34, pp. 39-58, 1991.
- [2] A. Følstad, E. Law, K. Hornbæk, "Analysis in practical usability evaluation: a survey study," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, 2012, pp. 2127-2136.
- [3] G. Kolfšchten, G.-J. D. Vreede, "The Collaboration Engineering Approach for Designing Collaboration Processes," in *International Conference on Groupware: Design, Implementation and Use*, 2007, pp. 38-54.
- [4] A. Solano, T. Granollers, C. Collazos, J. L. Arciniegas, "Experiencias en la especificación colaborativa de métodos de evaluación de usabilidad," in *XIV Congreso Internacional de Interacción Persona-Ordenador (Interacción 2013)*, Madrid, España, 2013.
- [5] D. Quiñones, C. Rusu, S. Roncagliolo, "Redefining Usability Heuristics for Transactional Web Applications," in *11th International Conference on Information Technology: New Generations*, 2014, pp. 260-265.
- [6] R. Otaiza, C. Rusu, S. Roncagliolo, "Evaluating the usability of transactional Web Sites," in *Third International Conference on Advances in Computer-Human Interactions (ACHI'10)*, Saint Maarten, 2010, pp. 32-37.
- [7] T. Hollingsed, D. G. Novick, "Usability inspection methods after 15 years of research and practice," in *Proceedings of the 25th annual ACM international conference on Design of communication*, 2007, pp. 249-255.
- [8] X. Ferré, "Marco de integración de la usabilidad en el proceso de desarrollo software," Tesis Doctoral, Lenguajes y Sistemas Informáticos e Ingeniería del Software, Universidad Politécnica de Madrid, Madrid, 2005.
- [9] T. Tullis, B. Albert, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*, Second ed.: Morgan Kaufmann, 2013.
- [10] M. Piattini, F. García, J. Garzàs, M. Genero, "Medición y estimación del software: técnicas y métodos para mejorar la calidad y productividad del software," *Ra-Ma (ed.)*, pp. 121-127, 2008.
- [11] D. C. Baird, J. C. Peña, *Experimentación: Una Introducción a la teoría de mediciones y al diseño de experimentos*: Prentice-Hall Hispanoamericana, 1991.