

MODELADO Y SIMULACIÓN DE LA EVALUACIÓN HEURÍSTICA DE USABILIDAD

Nuria Hurtado¹, Mercedes Ruiz¹ y Jesús Torres²

1: Departamento de Lenguajes y Sistemas Informáticos
Escuela Superior de Ingeniería
Universidad de Cádiz

C/ Chile nº1- 1003 – Cádiz (España)

e-mail: {nuria.hurtado, mercedes.ruiz}@uca.es, web: <http://www.uca.es>

2: Departamento de Lenguajes y Sistemas Informáticos
ETS Ingeniería Informática
Universidad de Sevilla

Avda. Reina Mercedes s/n - Sevilla (España)

e-mail: jtorres@lsi.us.es, web: <http://www.us.es>

Palabras clave: Modelado y Simulación del proceso software, Usabilidad, Evaluación Heurística.

Resumen. *La evaluación heurística o evaluación de expertos es, sin duda, el método de evaluación de usabilidad más conocido y utilizado. Dentro de los métodos de inspección de usabilidad, está considerado un método de bajo coste, especialmente indicado en circunstancias en las que los recursos y el presupuesto son limitados. El número de evaluadores que se utilice así como la experiencia de los mismos se consideran aspectos esenciales que afectan tanto a la calidad de las evaluaciones como a los costes que genera su aplicación. En este artículo, se presenta un modelo dinámico que simula el comportamiento del proceso de detección de problemas de usabilidad en una evaluación heurística. La simulación del modelo permite observar el comportamiento del método y la evolución en el tiempo de los costes asociados y la calidad del resultado en relación con el número de problemas de usabilidad encontrados. Del mismo modo, será posible predecir los efectos que tendrán en dicho comportamiento los cambios en factores clave como son el número de evaluadores y la experiencia de los mismos. Asimismo, el modelo desarrollado proporciona un marco para el estudio de otros factores que afectan a la evaluación heurística de usabilidad.*

1. INTRODUCCIÓN

Cada vez son más las compañías de desarrollo de software que incorporan aspectos de la ingeniería de la usabilidad a sus procesos de desarrollo. Sin embargo, es posible observar como hoy día la mayoría de las aplicaciones y sitios web todavía no proporcionan un adecuado nivel de usabilidad [2].

Uno de los factores que dificulta en gran medida la adopción de métodos de usabilidad es el elevado coste que supone la aplicación de complejas técnicas de evaluación. El uso de métodos de evaluación de usabilidad de bajo coste, como la evaluación heurística, puede contribuir a que una compañía aumente gradualmente la confianza en su eficacia y llegue a adoptar sistemáticamente métodos de usabilidad, comenzando desde lo más sencillo y progresando gradualmente hacia un ciclo de vida más refinado [19].

La evaluación heurística o evaluación de expertos es, sin duda, uno de los métodos de evaluación de usabilidad más utilizados [2]. Fue desarrollado originalmente por Nielsen y Molich [14][15] y consiste en analizar la conformidad de la interfaz con unos principios reconocidos de usabilidad denominados "heurísticas", mediante la inspección de varios evaluadores expertos. Es posible encontrar numerosas y completas referencias acerca de este método de evaluación y del proceso a seguir para llevarlo a cabo [3][4][12][18]. Pero sería interesante para los gestores y directores de proyectos disponer de una herramienta que les permitiera simular el comportamiento de dicho proceso y poder así predecir los efectos de determinados factores sobre áreas clave como la calidad, tiempo y coste de las evaluaciones, antes de realizarlas.

Las técnicas de modelado y simulación se han aplicado desde principios de los 90 para dar respuesta a distintas cuestiones relacionadas con el proceso software, considerándose herramientas valiosas en las tareas de gestión y de toma de decisiones. El objetivo de los modelos de simulación consiste en proporcionar mecanismos para la experimentación, predicción del comportamiento y aprendizaje del sistema representado [11].

En este artículo se presenta un modelo dinámico que simula el proceso de detección de problemas de usabilidad de una evaluación heurística. La simulación del modelo desarrollado permite predecir y valorar los efectos que tienen los cambios en factores clave, como son, el número y experiencia de los evaluadores, sobre la evolución del proceso de detección de problemas de usabilidad, concretamente, sobre el tiempo, coste y calidad de la evaluación.

En el apartado 2 se introducen los aspectos relativos al método de Evaluación Heurística de Usabilidad. El apartado 3 describe el desarrollo del modelo de simulación realizado, explicando detalladamente el propósito y ámbito del mismo, las variables y parámetros utilizados así como el proceso de abstracción seguido para su construcción. El apartado 4 presenta la simulación del modelo mostrando algunos resultados y análisis de diferentes experimentos. Finalmente se presentan las conclusiones obtenidas.

2. EVALUACIÓN HEURÍSTICA DE USABILIDAD

La Evaluación Heurística se engloba dentro de los denominados métodos de inspección de

usabilidad. Bajo esta clasificación se agrupan un conjunto de métodos en los que existen unos expertos que evalúan el grado de usabilidad de un sistema basándose en la inspección o examen de la interfaz del mismo. Una característica fundamental es que no existe presencia de usuarios durante la evaluación [5].

El método fue desarrollado originalmente por Nielsen y Molich [14][15] y consiste en analizar la conformidad de la interfaz con unos principios reconocidos de usabilidad denominados "heurísticas", mediante la inspección de varios evaluadores expertos.

Estas heurísticas son reglas generales que describen propiedades comunes de interfaces usables. Además de esta lista de heurísticas generales a considerar para todos los elementos de diálogo, es posible desarrollar heurísticas específicas que sean aplicables a un tipo concreto de productos como suplemento a las heurísticas generales.

Dentro de los métodos de inspección de usabilidad, la evaluación heurística es uno de los que implican más bajo coste. Diversas investigaciones demuestran que es un método muy eficiente y confirman su valor en circunstancias donde los recursos de tiempo y presupuesto disponibles son limitados [16].

El proceso para realizar una evaluación heurística se puede dividir fundamentalmente en tres fases: planificación, ejecución y por último, análisis y realización del informe final. En primer lugar se planifica la evaluación, eligiendo los evaluadores adecuados, los criterios que se adoptarán y se prepara el material necesario. La fase de ejecución es la más importante ya que es en la que se inspecciona la interfaz. Antes de comenzar las sesiones es posible realizar una sesión de formación con los evaluadores donde la persona que dirige la evaluación, denominada normalmente *observador*, explica las instrucciones sobre el método, los objetivos de la evaluación, los criterios heurísticos a seguir, así como los conceptos que sean necesarios sobre el dominio y propósito del sistema a evaluar.

Durante las sesiones de evaluación, el evaluador navega a través de la interfaz varias veces e inspecciona los elementos de diálogo, comparándolos con la lista de heurísticas establecidas. Es muy importante que las evaluaciones sean independientes para que no se produzcan desviaciones debido a las posibles influencias entre unos evaluadores y otros.

Finalmente cada evaluador proporciona un informe escrito con la lista de problemas encontrados con referencia a aquellas heurísticas que han sido violadas siguiendo algún tipo de plantilla previamente preparada.

Una vez que todas las evaluaciones han sido completadas estos problemas son analizados y agrupados obteniendo una única lista de problemas encontrados. La evaluación del grado de gravedad de los problemas se puede realizar durante las sesiones de evaluación y reflejarla en los informes individuales que posteriormente se analizarán, o bien, una vez que los problemas se agrupan, se lleva a cabo una puesta en común sobre la evaluación y posteriormente los evaluadores califican la gravedad de los problemas finales.

Otro modo de conducir las sesiones de evaluación, es contar con la figura del observador al que se le irán comunicando los problemas encontrados durante la evaluación y que se encargará de ir recopilando y agrupando la información para generar un único informe. Los informes escritos individuales tienen la ventaja de proporcionar un registro formal de la evaluación completa, pero requieren un esfuerzo adicional por parte de los evaluadores,

así como por parte del observador que posteriormente se encargará de leerlos y analizarlos. Utilizar un observador, incrementa los costes generales del proceso de detección, pero tiene la ventaja de reducir la cantidad de trabajo posterior, ya que la agregación de los diferentes problemas encontrados se realiza en paralelo con las sesiones de evaluación y los resultados finales estarán disponibles más rápidamente [18].

Son varios los factores que pueden influir en el desarrollo y evolución del método. En este estudio se considerarán algunos de los más importantes como son el número y experiencia de los evaluadores.

2.1. Número y experiencia de los evaluadores

En principio, es posible realizar evaluaciones de usabilidad basadas en un solo evaluador, pero las referencias estudiadas concluyen que los resultados no son satisfactorios, ya que una sola persona no es capaz de encontrar todos los problemas de usabilidad de la interfaz. Referencias relativas a distintos proyectos demuestran que diferentes personas encuentran diferentes problemas de usabilidad, esto se denomina *efecto evaluador* [6][7][18]. Es difícil encontrar al mejor evaluador y usarlo para todas las evaluaciones ya que la misma persona no es siempre la mejor en todas las evaluaciones, incluso algunos de los problemas más difíciles de encontrar son a veces encontrados por evaluadores que no encuentran la mayoría de los problemas, de este modo es importante implicar múltiples evaluadores en cualquier evaluación heurística [15][18].

Nielsen y Landauer plantean la detección de problemas de usabilidad como función del número de evaluadores usando un proceso Poisson, representado en la ecuación (1) [17].

$$problemas_encontrados(i) = N(1 - (1 - \lambda)^i) \quad (1)$$

Donde *problemas_encontrados(i)* indica el número de problemas de usabilidad diferentes encontrados tras la agregación de informes de *i* evaluadores independientes, *N* representa el número total de problemas de la interfaz y λ la proporción de problemas encontrados por un único evaluador, valores que cambiarán dependiendo de las características del proyecto.

Aunque la evaluación heurística fue originalmente desarrollada como un método de ingeniería de la usabilidad para evaluadores que tenían alguna experiencia en usabilidad pero que no eran necesariamente expertos, diferentes estudios demuestran que la experiencia de los evaluadores mejora considerablemente la eficacia del método [16].

3. DESARROLLO DEL MODELO DE SIMULACIÓN

Desde principios de los 90 se han desarrollado diferentes modelos de simulación aplicados al proceso de desarrollo de software, resaltando especialmente los modelos de Abdel-Hamid [1] y Madachy [13]. Se han realizado trabajos previos en el ámbito del modelado y simulación aplicado a la mejora de la usabilidad, concretamente en el desarrollo de modelos de simulación que representan la integración del proceso de desarrollo centrado

en el usuario (DCU) en el ciclo de vida de desarrollo de software [8][9][10]. En los siguientes apartados se describe detalladamente el proceso de realización del modelo propuesto en el presente trabajo.

3.1. Propósito y ámbito del modelo

El modelo desarrollado tiene como objetivo servir de herramienta de ayuda a la toma de decisiones, ya que será posible predecir los efectos que tienen los cambios en el número y experiencia de los evaluadores sobre la evolución en el tiempo de la evaluación heurística. El ámbito del modelo en este caso, abarca la fase de detección y análisis de problemas de la evaluación heurística de usabilidad que da como resultado un único conjunto de problemas de usabilidad encontrados. Para realizar el modelo se ha supuesto que durante las sesiones de evaluación se realiza también el análisis y agregación de los problemas encontrados en las evaluaciones independientes, es decir, se analizan y eliminan los solapamientos entre los problemas encontrados por diferentes evaluadores y se va construyendo una única lista de problemas de usabilidad final. Se ha elegido exclusivamente esta fase dependiente de los evaluadores ya que son precisamente los efectos de la variación en este factor los que se pretenden simular con el modelo desarrollado, considerando además que se trata de la fase esencial del método.

3.2. Parámetros de entrada y variables resultado

Los principales **parámetros de entrada** al modelo son:

- *Nº evaluadores $E(i)$* : Este parámetro se subdivide en otros tres parámetros dependiendo del nivel de experiencia de modo que sea posible realizar simulaciones en las que se mezclen evaluadores con diferente nivel de experiencia. Se han considerado tres niveles de experiencia extraídos de [16]. Nº Evaluadores E1 representa el número de evaluadores noveles o con escasa experiencia en usabilidad. Nº Evaluadores E2 representa el número de evaluadores expertos en usabilidad. Nº Evaluadores E3 representa el número de evaluadores expertos en usabilidad así como en el dominio concreto del sistema a evaluar.
- *Problemas de usabilidad de la interfaz*: representa el número de problemas totales de usabilidad estimados.
- *Capacidad $E(i)$* : Este parámetro se subdivide también en E1, E2 y E3 dependiendo del nivel de experiencia y representará la cantidad de problemas de usabilidad que es capaz de resolver un evaluador por unidad de tiempo (hora).
- *Coste $E(i)$* : Este parámetro se subdivide también en E1, E2 y E3 dependiendo del nivel de experiencia y representará el coste (euros) por unidad de tiempo (hora) que representa cada evaluador, incluye el sueldo del evaluador así como el coste de analizar su informe y el coste de cualquier ordenador o recurso que se utilice durante la evaluación [18].
- *Proporción $E(i)$* : Este parámetro se subdivide también en E1, E2 y E3 dependiendo del nivel de experiencia y representará la proporción media de problemas que es capaz de

encontrar un solo evaluador. Podrá ser estimado u obtenido de otros estudios similares, a partir de la recogida de métricas de proyectos anteriores, o bien a partir del modelo matemático de Nielsen y Landauer [17].

Las principales **variables resultado** del modelo son:

- *Problemas de usabilidad encontrados*: representa el número de problemas de usabilidad diferentes encontrados tras el análisis y agregación de las evaluaciones independientes, determinará por tanto, la calidad de la evaluación.
- *Coste variable acumulado*: representa el coste variable que implica la aplicación del método, aquellos que se incrementan cada vez que se usa un evaluador adicional.

3.3. Proceso de abstracción del modelo

La figura 1 muestra el diagrama de flujos y niveles simplificado del modelo, en el que se pueden observar, además de las comentadas anteriormente, las variables auxiliares que se han utilizado para representar el proceso. Para simular el proceso de detección y análisis de problemas de usabilidad se han supuesto tres variables de nivel: *Problemas de usabilidad por detectar* que se inicializará con el tamaño estimado de problemas de usabilidad que tiene la interfaz, *Problemas de usabilidad encontrados* y *coste variable acumulado*. El flujo de problemas transcurre de *Problemas por detectar* a *problemas detectados* a través de una variable de flujo denominada *tasa de detección de problemas*. Esta tasa irá aplicando progresivamente la *capacidad real* de detección según los parámetros de entrada hasta que la cantidad de problemas que queden por detectar sea menor que la *capacidad real*, entonces se aplicará la *capacidad necesaria* para terminar y el flujo de detección se detendrá.

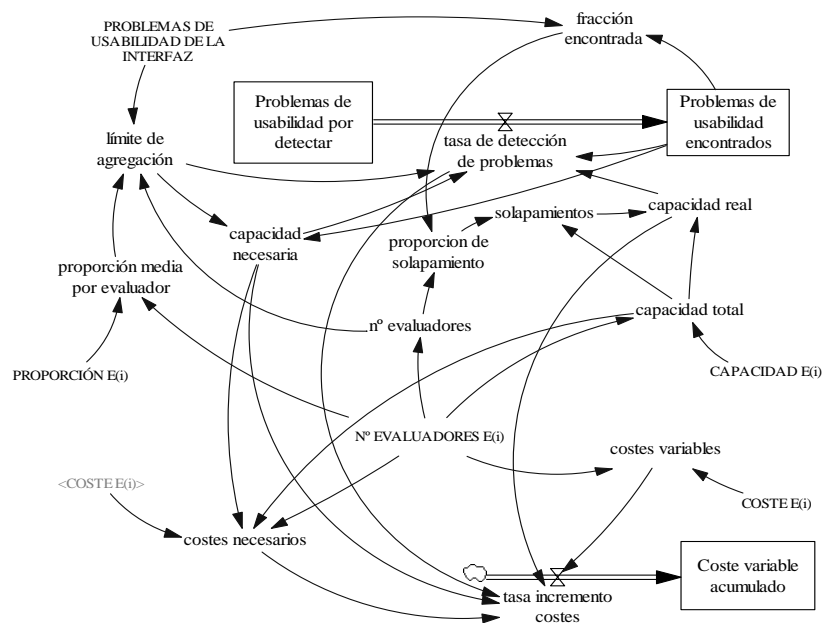


Figura 1. Diagrama de flujos y niveles simplificado del modelo

Conforme se realiza la evaluación, los problemas van siendo analizados por un observador que elimina los solapamientos y construye una única lista de problemas finales. Esto se ha simulado suponiendo una *proporción de solapamiento* que se ha considerado dependiente del número de evaluadores y de la fracción de problemas encontrada, teniendo en cuenta que entre dos evaluadores cualesquiera existe un solapamiento medio del 9% [7]. El *límite de agregación* determina la cantidad de problemas que es posible resolver según sea el número de evaluadores y su experiencia, para ello se ha utilizado la ecuación (1), que servirá para detener la tasa de detección de problemas. Paralelamente, a medida que avanza la evaluación la *tasa de incremento de costes* fluye hacia la variable de nivel *coste variable acumulado*, incrementando estos costes según sea el número, experiencia y coste asociado a cada nivel de experiencia. Este flujo se detendrá cuando finalice la evaluación.

4. SIMULACIÓN DEL MODELO

Una vez introducidas todas las ecuaciones que gobiernan el comportamiento del modelo se puede proceder a su simulación realizando tantos experimentos como se desee. Considerando que el propósito del modelo es observar las influencias de las variaciones en el número y experiencia de los evaluadores, se fijan el resto de los parámetros, que dependerán del experimento que se desee realizar:

- Problemas de usabilidad de la interfaz: 100.
- Proporciones E1, E2, E3: 0.2, 0.3 y 0.4, respectivamente. Extraídos de [16].
- Capacidades E1, E2, E3: se han supuesto las constantes 2, 4 y 6, respectivamente.
- Costes E1, E2, E3: 50, 100 y 150, respectivamente.

4.1.- Resultados de la simulación

En la figura 2 se observa la evolución del porcentaje de problemas de usabilidad encontrados simulando diferentes escenarios según el nivel de experiencia y variando el número de evaluadores. Se puede observar que los resultados siguen tendencias similares a los modelos de referencia utilizados [16][18].

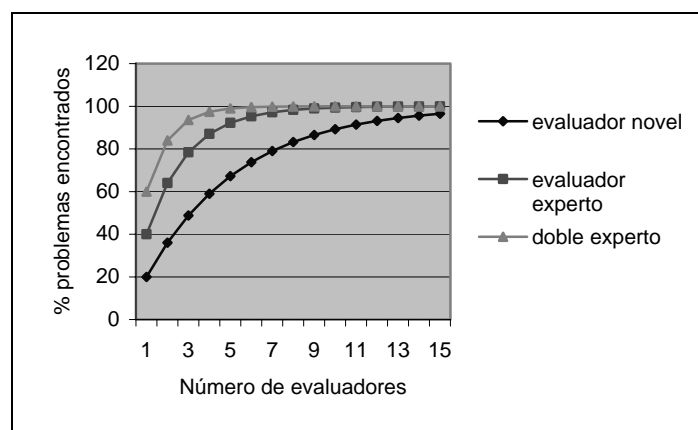


Figura 2. Proporción de problemas encontrados según el número de evaluadores.

La figura 3 muestra las gráficas de la evolución en el tiempo de las variables resultado *problemas de usabilidad encontrados* y *coste variable acumulado* para un nivel de experiencia 1. Cada línea de estas gráficas representa una simulación del modelo, estas se han nombrado como *nivel de experiencia_ número de evaluadores*.

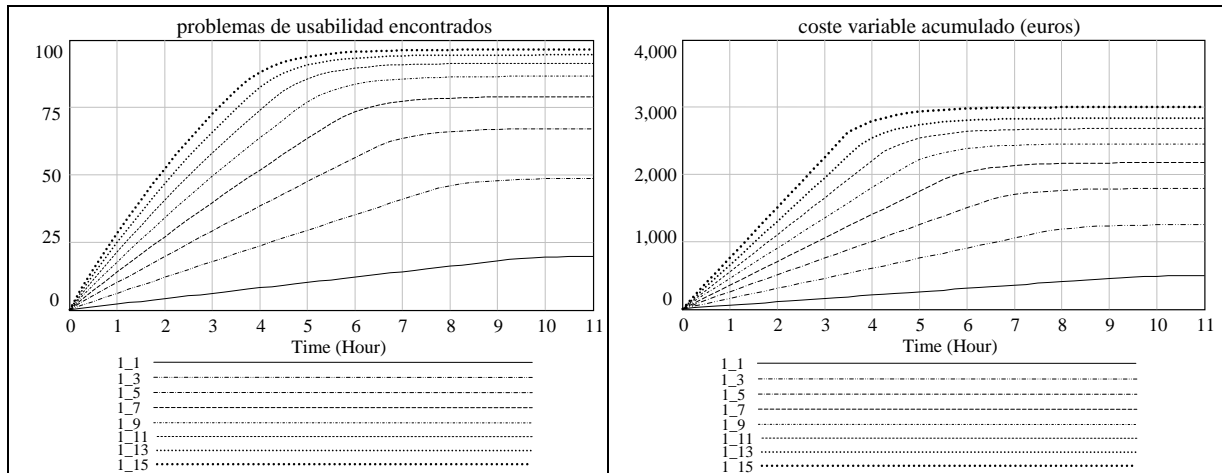


Figura 3. Comparación usando diferente número de evaluadores noveles.

Observando este tipo de resultados es posible estimar por ejemplo el porcentaje de problemas que se ha resuelto en un momento determinado, el tiempo que se tarda en finalizar la evaluación o el tiempo y coste necesarios para conseguir un determinado resultado, de este modo es posible tomar decisiones relativas a cuándo interesa parar de evaluar o que combinación de evaluadores es la más conveniente, dependiendo de los recursos de la organización, de lo crucial que sea la usabilidad del sistema y del beneficio esperado de mejorar la usabilidad. La figura 4 muestra el comportamiento de las variables resultado variando el nivel de experiencia y utilizando un valor fijo de 5 evaluadores.

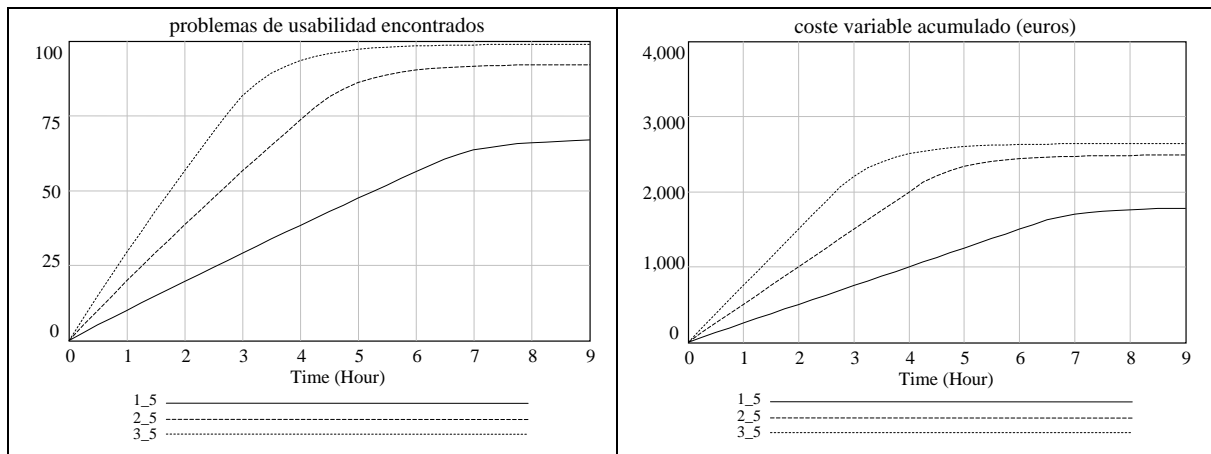


Figura 4. Comparación de niveles de experiencia usando 5 evaluadores.

5. CONCLUSIONES

Como principal conclusión, este trabajo ilustra la aportación original que las técnicas de modelado y simulación pueden suponer aplicadas a los métodos de evaluación de la usabilidad. En este artículo se ha presentado el desarrollo de un modelo que permite simular la evolución en el tiempo del proceso de detección y análisis de problemas de una evaluación heurística. Para ello se ha tenido en cuenta fundamentalmente el modelo matemático de Nielsen y Landauer y sus conclusiones respecto a la influencia del número y experiencia de los evaluadores en la proporción de problemas que es posible detectar [16][17]. La aplicación de sistemas dinámicos hace posible añadir a estas conclusiones, el estudio de la evolución en el tiempo del proceso y como consecuencia del coste asociado al mismo, factores esenciales para cualquier organización de desarrollo de software que pretenda aplicar algún método de evaluación de la usabilidad. El modelo presentado permite también observar el comportamiento del método cuando se mezclan en una misma evaluación evaluadores con diferente experiencia.

Por todo esto, este modelo proporciona una herramienta de utilidad para la experimentación de diferentes políticas que puede servir de ayuda a la toma de decisiones en la aplicación de la evaluación heurística de usabilidad. Del mismo modo, el modelo desarrollado se plantea como un marco de estudio sobre el que será posible añadir otros factores que afectan al comportamiento del método como pueden ser: el tipo de prototipo que se use, la complejidad de los problemas o las heurísticas que se utilicen para realizar la evaluación.

AGRADECIMIENTOS

Los autores agradecen a la Comisión Interministerial de Ciencia y Tecnología la subvención de esta investigación a través de los proyectos TIC 2003-369 y TIN2004-06689-C03-03.

REFERENCIAS

- [1] T. Abdel-Hamid and S. Madnick, *Software Project Dynamics: An Integrated Approach*. Prentice-Hall, Englewood Cliffs, NJ.(1991).
- [2] R.G. Bias and D.J. Mayhew (eds). *Cost-Justifying Usability. An update for the internet age*, Elsevier, (2005).
- [3] O. Daly-Jones, N. Bevan and C. Thomas, *Handbook of User Centred Design*. IE2016 INUSE Deliverable D6.2.1 Version 1.31. Serco Usability Services, National Physical Laboratory, Teddington, Middx, UK. (2001).
- [4] J.S. Dumas and J.C. Redish, *A practical guide to usability testing*. Rev. ed. Exeter, England ; Portland, Or. : Intellect Books, c1999, (1999).
- [5] T. Granollers, *MPLu+a, Una Metodología que integra la Ingeniería del Software, La Interacción Persona Ordenador y la accesibilidad en el Contexto de equipos de desarrollo Multidisciplinares*. Tesis Doctoral. Departamento de Lenguajes y Sistemas Informáticos, Universidad de Lleida, (2004).

- [6] M. Hertzum and N.E. Jacobsen, *The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods*. Proceedings of *International Journal of Human-Computer Interaction*, vol. 13, no. 4, pp. 421-443. (2001).
- [7] M. Hertzum, N.E. Jacobsen and R. Molich, *Usability Inspections by Groups of Specialists: Perceived Agreement in Spite of Disparate Observations*. Proceedings of *Conference on Human Factors in Computing Systems, CHI '02*, pp 662-663, (2002).
- [8] N. Hurtado, M. Ruiz, J. Torres, *Aplicación del Modelado y Simulación de Sistemas Dinámicos al Proceso de Diseño Centrado en el Usuario*. Actas de ADIS'04. JISBD 2004. Universidad de Málaga. Noviembre (2004).
- [9] N. Hurtado, M. Ruiz, J. Torres, *Modelado y simulación como herramienta para la mejora de la usabilidad*. Actas de *interacción 2005*. Celebrado junto con el *I Congreso Español de Informática*, CEDI 2005, Granada, Septiembre (2005).
- [10] N. Hurtado, M. Ruiz and J. Torres, *Towards Interactive System Usability through Simulation Modeling*. Proceedings of the *6th International Workshop on Software Process Simulation and Modeling, ProSim 2005*. Co-located with ICSE 2005. St. Louis, MO, USA, (2005).
- [11] M.I. Kellner, R.J. Madachy and D.M. Raffo, *Software Process Simulation Modeling: Why? What? How?. The Journal of Systems and Software*, 46 (2/3), pp. 91-105, (1999).
- [12] G. Lindgaard, *Usability testing and system evaluation*. London: Chapman & Hall. (1994).
- [13] R. Madachy, *A Software Project Dynamics Model for Process Cost, Schedule and Risk Assessment*. Ph.D. Dissertation. University of Southern California, Los Angeles, CA. (1994).
- [14] R. Molich, and J. Nielsen, *Improving a human-computer dialogue*. Communications of the ACM, vol. 33, Issue 3, (1990).
- [15] J. Nielsen and R., Molich, *Heuristic evaluation of user interfaces*. Proceedings of the *SIGCHI conference on human factors in computing systems*. Seattle, WA, USA, pp 249-256. (1990).
- [16] J. Nielsen, *Finding usability problems through heuristic evaluation*. Proceedings of the *SIGCHI conference on human factors in computing systems, ACM CHI'92*. Monterey, CA, USA, pp 373-380. (1992).
- [17] J. Nielsen and T.K. Landauer, *A mathematical model of the finding of usability problems*. Proceedings of the *ACM INTERCHI'93 Conference*. Amsterdam, the Netherlands, pp 206-213, (1993).
- [18] J. Nielsen, *Heuristic evaluation*. In J. Nielsen, and R.L. Mack, (eds), *Usability Inspection Methods*. John Wiley & Sons, New York, NY. (1994).
- [19] J. Nielsen. *Guerrilla HCI: Using Discount Usability Engineering to penetrate the intimidation barrier*. In R. G. Bias, and D. J. Mayhew (Eds.), *Cost-Justifying Usability*. Boston, MA: Academic Press, pp 245-272, (1994).