

Guía de Evaluación de Usabilidad de Software

Nelson Baloian y José A. Pino

Marzo de 2024

DCC – Universidad de Chile

1. Introducción

La usabilidad de una aplicación de software tiene relación con la facilidad con la cual otras personas pueden utilizar esa aplicación. Esta facilidad tiene componentes subjetivos, tales como la preparación previa del usuario, sus gustos, incluso su historia personal.

Es así como un determinado software podría ser considerado como “muy usable” por una persona, pero no así por otras personas. Quizás, incluso podrían buscarse grupos o “poblaciones” para las cuales un software sería muy usable. Ejemplo de esto podría ser “niñas de 10 a 12 años”, otro podría ser “adultos mayores”, y un tercero, “personas con discapacidad manual”.

En todo caso, existen principios que orientan sobre lo que es considerado deseable por la gran mayoría de las personas, y que se pueden consultar en algún libro de Interacción Humano-Computador. Por ejemplo, a las personas les gusta la simetría en una figura, o los rectángulos cerrados.

La usabilidad se asocia a múltiples dimensiones. En particular, tradicionalmente se la asocia a cinco atributos:

- a) Facilidad de aprendizaje.
- b) Eficiencia. Una vez que se ha aprendido a usar, el software debiera permitir alto nivel de productividad.
- c) Facilidad de recordar. El software debiera ser fácil de recordar: si el usuario vuelve a usar el software después de un período de no utilización, no debiera tener que aprender todo de nuevo.
- d) Errores. El usuario debiera cometer pocos errores usando el sistema. Si los comete, debiera poder recuperarse fácilmente de ellos. No debieran ocurrir errores catastróficos.
- e) Satisfacción. El software debe ser agradable de usar.

Esta guía no pretende cubrir todo lo referente a la usabilidad. Sólo se refiere, y en forma resumida, principalmente, a satisfacer una pregunta recurrente para el desarrollador típico de software: “tengo un software ya desarrollado, ¿es usable por mis usuarios objetivo?”. Naturalmente, la respuesta a esta pregunta requiere hacer algún trabajo, que se explica en esta guía.

Antes de proceder a trabajar la respuesta apropiada, se mencionan a continuación algunas respuestas **incorrectas** a la pregunta:

- i. *A mí me gustó como quedó, por lo tanto, a otros también les va a gustar.* Curiosamente, esta respuesta es la que inocentemente muchos desarrolladores dan. Lo que me gusta a mí no necesariamente es apropiado para los usuarios objetivo: normalmente **no** lo es.
- ii. *Se lo pasé a la persona X, lo probó, y le pareció muy apropiado.* X no necesariamente representa a los usuarios objetivo.

- iii. *El software tiene una interfaz igual al del software Y, que se usa ampliamente.* El software desarrollado no necesariamente hace lo mismo que el software Y.
- iv. *No tengo acceso a usuarios que puedan probar el software; en consecuencia, no puedo decir nada respecto a la usabilidad.* Es posible afirmar algo sobre la usabilidad de un software sin recurrir a usuarios, como veremos en esta guía.
- v. *No tengo tiempo suficiente para hacer evaluaciones de usabilidad.* Existen métodos rápidos de evaluación de usabilidad.

El resto de esta guía se organiza como sigue: El Capítulo 2 presenta los principios de usabilidad. El capítulo 3 clasifica los métodos de evaluación de usabilidad. Los siguientes capítulos discuten métodos específicos de evaluación. Si quiere usar algún cuestionario, revisamos tres de los más conocidos, aunque, por supuesto, existen otros; en este caso, recomendamos leer primero el capítulo 10, que hace consideraciones generales, pero importantes, sobre el uso de cuestionarios.

Esta guía está extraída de parte de los apuntes de nuestro curso CC 5511 “Fundamentos e Ingeniería de Interacciones Human-Computador” del Departamento de Ciencias de la Computación de la Universidad de Chile.

2. Principios generales de diseño de interacción

Antes de estudiar los métodos de evaluación de interfaces humano-computador, es conveniente detenernos en ciertas ideas o principios básicos sobre diseño de interfaces.

A continuación, se presentan diez principios propuestos en 1994 por Jakob Nielsen [1] con el nombre de “heurísticas de usabilidad”:

1. **Visibilidad del estado del sistema**
El diseño del sistema debe mantener siempre informado al usuario de lo que está pasando, con retroalimentación apropiada, dentro de un plazo razonable.
2. **Coincidencia del sistema con el mundo real**
El diseño debe utilizar el lenguaje del usuario. Utilice palabras, frases y conceptos familiares al usuario, en vez de terminología interna (computacional). Siga las convenciones del mundo real, haciendo que la información aparezca en un orden natural y lógico.
3. **Control y libertad del usuario**
Los usuarios a menudo realizan acciones por error. Necesitan una “salida de emergencia” claramente rotulada para abandonar la acción no deseada sin tener que pasar por un largo proceso.
4. **Consistencia y estándares**
Los usuarios no tienen que dilucidar si diferentes palabras, situaciones o acciones son la misma cosa o no. Utilice las convenciones de la industria y de la plataforma usada.
5. **Prevención de errores**
Tener buenos mensajes de error es importante, pero los mejores diseños evitan que se produzcan errores en primer lugar. Ya sea elimine las situaciones propensas a error, o chequéelas y pida a los usuarios una confirmación antes de que efectúen una acción.
6. **Reconocer antes que recordar**
Minimice la carga sobre la memoria del usuario mostrando los elementos, acciones y opciones. El usuario no debería tener que recordar información al pasar desde una parte de la interfaz a otra. La información requerida para usar el diseño (por ej., rótulos de campos o ítems de menú) deben ser visibles o fácilmente recuperables.

7. Flexibilidad y eficiencia de uso

Los atajos – escondidos para los usuarios novicios – pueden acelerar la interacción para los usuarios expertos, de modo que el diseño sirva tanto para los usuarios sin experiencia previa, como para los experimentados. Permita que los usuarios ajusten las acciones frecuentes.

8. Diseño estético y minimalista

Las interfaces no debieran incluir elementos irrelevantes o utilizados extremadamente poco. Cada unidad adicional de información compite con las unidades relevantes de información y disminuye su visibilidad relativa.

9. Ayude a los usuarios a reconocer, diagnosticar y recuperarse de errores

Los mensajes de error deben expresarse en lenguaje corriente (no códigos de error). Además, deben indicar el problema con precisión, y sugerir constructivamente una solución.

10. Ayuda y documentación

Lo mejor es que el sistema no necesite ninguna explicación adicional. Sin embargo, puede ser necesario proveer documentación para ayudar a los usuarios a entender cómo completar sus tareas.

Note que estos principios cubren las dimensiones de usabilidad que se mencionan en la Introducción. Esto previene una inclinación inicial a concentrarse en sólo una dimensión e ignorar otras. Por ejemplo, concentrarse sólo en el diseño estético e ignorar el tema de los errores, en circunstancias que los usuarios reales con certeza cometerán errores.

3. Clasificando métodos de evaluación

Existen numerosos métodos de evaluación, y se pueden inventar nuevos. También se les puede combinar, para aprovechar las ventajas que puedan tener dos o más métodos.

En los siguientes sub-capítulos, clasificaremos los métodos de acuerdo a varios criterios.

3.1. Rigurosidad

La evaluación de la usabilidad de un software puede necesitarse para múltiples propósitos. Por ejemplo, puede requerirse para tener una primera aproximación a la usabilidad del software. En contraste, podría usarse para justificar en un artículo científico la usabilidad de un software desarrollado. Note que estos propósitos son muy distintos. Uno podría equivocarse en una primera apreciación, pero no en un artículo científico.

En el primer caso (primera aproximación) puede usarse un método que llamaremos **rápido**, porque seguramente tiene bajo costo y se hace en poco tiempo, para hacer una estimación que sea suficiente para la situación actual. En el segundo caso (certeza), hay que usar un método que llamaremos **riguroso**.

Note que no es conveniente confundir esta clasificación básica. Nos ha correspondido evaluar tesis en las cuales se usa un método rápido, y al final, en las Conclusiones, el tesista hace afirmaciones del tipo “Queda demostrada la usabilidad de la aplicación...”. En tales casos, el tesista debe cambiar la redacción (si es que es posible) para afirmar, que “hay indicios de una buena usabilidad”, o bien, cambiar el método de evaluación.

Veamos las características de estos dos tipos de evaluación.

Un método de evaluación **riguroso** debe ser simultáneamente:

- **Preciso:** el método debe ser capaz de medir adecuadamente las variables evaluadas, dentro de un cierto margen de error.
- **Confiable:** el método establece una relación consistente entre las variables y sus valores a través de varias evaluaciones.
- **Generalizable:** el método provee medidas que pueden ser usadas con confianza para diferentes casos y sesiones de evaluación.

Las tres propiedades son necesarias para poder decir “diseño A es mejor que diseño B”. Sin método preciso, uno nunca puede estar seguro si las diferencias medidas entre dos diseños son debidas al cambio de diseño o no. Sin método confiable,

podemos obtener resultados que no son comparables con otros, o con respecto a una línea base. Sin generalizabilidad, los resultados no son válidos más allá del proyecto de evaluación.

Los métodos que satisfacen estas tres propiedades han sido probados ampliamente. De este modo, podemos decir que son robustos. También, estos métodos son considerados buenas prácticas.

El problema con los métodos rigurosos es que cumplir con las propiedades involucra típicamente un alto costo. La persona que realiza la evaluación, para empezar, debe saber bien las restricciones y requerimientos del método seleccionado. También debe saber los protocolos y reglas del método, que pueden ser complejos de implementar y pueden restringir severamente las condiciones en las cuales se hacen las sesiones de evaluación. Las sesiones de evaluación pueden demorar mucho, repetirse múltiples veces, y requerir alto número de participantes.

Ejemplo: En una universidad se desarrolló un sistema de apoyo a toma de decisiones de crowdsourcing. Teniendo el resultado, se quería publicar un estudio científico sobre el valor del sistema.



Propósito de la evaluación: estudiar desempeño humano con y sin sistema.

Planificación cuidadosa: se hizo un trabajo para definición de variables y medidas, definición de ejercicios, selección de participantes.

Se hicieron sesiones experimentales con 60 participantes.

Los resultados mostraron que los ejercicios habían sido demasiado simples, y los resultados eran inútiles para ser significativos. Se tuvo que hacer un nuevo experimento, con nuevos ejercicios y 200 participantes. Esta vez sí se tuvo resultados apropiados, aunque a un alto costo. ■

El ejemplo muestra un caso típico de lo costoso que puede ser un método riguroso. ¿Lo necesitamos realmente? En muchos casos, basta con saber que el diseño es razonablemente bueno, o que va en la dirección correcta.

En otros casos, hay que proveer certezas, requerido por ley, práctica profesional o estándar industrial. En industrias tales como la aviación, control de tráfico aéreo, salud, se requiere evidencia específica de calidad, demostrada rigurosamente.

Una evaluación **rápida** es lo contrario de la rigurosa. No hay requerimiento específico (excepto que sea barato). No se necesita ni espera precisión, rigor, ni generalizabilidad. Los protocolos son simples y flexibles. Cada sesión puede ajustarse a una necesidad particular o para seguir un evento específico. Por otro lado, no se puede hacer afirmaciones sólidas acerca de los resultados. Los resultados son para usarse internamente en algún proyecto.

Ejemplo: desarrollo de herramienta de asignación de objetivos en móviles. La herramienta pedía tareas que incluían recolección de datos.

Objetivo: evaluar el mecanismo de recolección de datos



El método adoptado consistía en registrar datos y usar una cámara para obtener información contextual, acerca de dónde se estaba moviendo el usuario. Se daban instrucciones sueltas a los usuarios, y se cambiaban de usuario a usuario para enfocarse en eventos específicos que se consideraban de interés para el diseñador.

■

En este ejemplo, vemos cómo un método rápido permitía hacer cambios sobre la marcha, para satisfacer la tarea de cómo desarrollar mejor el sistema.

Finalmente, la siguiente tabla resume las diferencias entre métodos rigurosos y rápidos.

Evaluación rigurosa	Evaluación rápida
Proceso Formal	No hay reglas en particular
Método conocido, aceptado por la comunidad	Método puede ser conocido o desconocido
El foco está en precisión, confiabilidad y generalizabilidad	El foco está en retroalimentación rápida
Alta calidad de los datos	Baja calidad de los datos

3.2 Evaluación Sumativa o Formativa

Un segundo criterio de clasificación de métodos de evaluación se refiere al momento en que se hace la evaluación.

Una evaluación sumativa se hace al final de un proyecto de desarrollo. Se quiere dar una conclusión sobre el resultado. Por el contrario, una evaluación formativa se hace durante el desarrollo y sirve para orientar el resto del desarrollo, o corregir deficiencias que se detecten. Note que puede haber varias evaluaciones formativas a lo largo de un desarrollo de software.

En una evaluación sumativa, quienes la realizan deben ser externos al equipo de desarrollo del software. Así se asegura que no haya sesgos internos.

Evaluación sumativa	Evaluación formativa
Al final del proyecto, paso final en el proceso de desarrollo	Durante el proyecto, parte del ciclo de desarrollo
Se hace una vez	Se hace múltiples veces
Artefacto completo, funcional	Artefactos intermedios
Idealmente, externa	Usualmente interna

3.3 Evaluación Formal o de Descuento

El concepto de evaluación de descuento (“discount method”) se refiere a un método efectivo, pero de bajo costo. Inicialmente se aplicaba a tres métodos específicos,

pero después el término se ha aplicado a otros métodos baratos. En contraste, un método “formal” implica un alto costo de aplicación.

3.4 Evaluación Cuantitativa versus Cualitativa

La evaluación cuantitativa se enfoca en describir un fenómeno usando cantidades o magnitudes de variación. El proceso evaluativo necesariamente tiene que ver con definición precisa de variables, escalas, y procedimientos para obtener las medidas. La evaluación cuantitativa se asocia a una visión **positivista** de los fenómenos. Esto quiere decir que si existe un fenómeno en el mundo real, entonces podemos encontrar su evidencia empírica, la que puede ser medida objetivamente.

Ejemplos de evaluación cuantitativa son

- Eficiencia: número de clicks para realizar tarea, tiempo de completación, tiempo de respuesta del sistema, etc.
- Efectividad: proporción de éxito, tiempo de aprendizaje, número de errores de usuario, etc.
- Retroalimentación de usuarios: satisfacción, facilidad de uso, facilidad de aprendizaje, probabilidad de regreso, rapidez percibida, frecuencia de uso, etc.
- Diseños web: Número de visitas, Page views, Duración de las visitas.

La evaluación cualitativa, por el contrario, se enfoca en describir un fenómeno sin cuantificarlo. Por ejemplo, podemos describir los componentes de un fenómeno, o podemos caracterizar un fenómeno usando escalas ordinales o nominales. Las escalas ordinales incorporan orden, pero no “distancia” entre valores. Las escalas nominales representan categorías de valores, pero no orden entre ellos.

Ejemplos de evaluación cualitativa son:

- Ejecución de tareas: proceso de pensar, proceso de aprender, flujo de interacciones.
- Comportamiento de usuarios: movimiento de ojos, comportamiento de búsqueda, dificultades de usuario, atajos, etc.
- Percepción de usuarios: comprensión de conceptos, estructura percibida, seguridad, utilidad, satisfacción logrando objetivos, etc.
- En diseño web: Seguimiento de clicks, mapas de clicks, mapas de calor, mapas de scroll, búsquedas.

Evaluación cuantitativa	Evaluación cualitativa
Basada en medidas	Basada en conceptos e indicadores
Usa proporciones, intervalos y escalas absolutas	Usa escalas nominales u ordinales
El foco está en cuantificación	El foco está en comprensión
Describe un fenómeno usando cantidades y magnitud de variaciones	Describe un fenómeno sin cuantificarlo

3.5 Evaluación con usuarios o evaluación sin usuarios

Una propuesta natural es recurrir a futuros usuarios de un software para que evalúen el nuevo producto. En unos pocos casos, puede que esos usuarios estén disponibles y deseosos de probar el nuevo software. Sería buena cosa. Sin embargo, en la inmensa mayoría de los casos, esos usuarios no están disponibles o lo están a muy alto costo (medido de alguna manera). Por ejemplo, imaginemos un nuevo software a utilizarse en la mesa de dinero de un banco; normalmente, los operadores están demasiado ocupados y sobrecargados de trabajo como para destinar tiempo a evaluar un nuevo software; aquí tenemos un caso de usuarios que simplemente no están disponibles.

Afortunadamente, hay métodos de evaluación que no requieren la presencia de usuarios. Se basan en estudiar características del software mismo y pueden ser muy apropiados en ciertos casos. Algunos de estos métodos son rigurosos.

Volvamos al caso de la evaluación con usuarios. El ideal es que los usuarios sean exactamente las mismas personas que posteriormente van a utilizar el software (como en el ejemplo anterior de los operadores de mesa de dinero). Pero hay situaciones en las que podemos encontrar “representantes” de los usuarios. Por ejemplo, supongamos un software para niñas entre 11 y 13 años. Resultaría imposible conseguir a todas las niñas en ese rango etario. Podemos entonces buscar un cierto número de niñas que representen al universo de las niñas entre 11 y 13 años. En estos casos de búsqueda de usuarios representativos, deben tenerse en cuenta los siguientes criterios básicos:

- Independencia: los usuarios deben tener independencia del desarrollador: no ser sus familiares, por ejemplo.
- Cuidado con los incentivos: si se da incentivos a los usuarios (por ej., monetarios), éstos deben informarse de tal modo que los usuarios entiendan bien que recibirán el incentivo independientemente de cómo evalúen el software, pero que deben hacerlo a conciencia.

- Número: debe haber un número mínimo de usuarios. Esto depende del método de evaluación. Como regla general, se puede pensar en un mínimo de 20 usuarios. Hay casos de evaluación rigurosa en que se requieren 200 usuarios o más.
- Representatividad: los usuarios que van a evaluar deben tener idealmente una estrecha relación con los usuarios finales del software. Por ejemplo, en el caso del software para mesa de dinero, se podría utilizar operadores “trainees”, es decir, no operadores verdaderos, pero personas que ya han sido entrenados como operadores, aunque todavía no trabajan como tales.
- Un tema especial es la utilización de estudiantes universitarios como usuarios. En muchos casos en que resulta difícil o imposible conseguir usuarios reales, los desarrolladores tienden a utilizar estudiantes universitarios como usuarios. Hasta hace unos años, los journals de Sistemas de Información rechazaban artículos científicos en que se presentaba como evaluación rigurosa evaluaciones con estudiantes como usuarios (de software no educacional). Sin embargo, en años recientes, se ha flexibilizado esa regla, y ahora, en general, se acepta, pero, por otro lado, se exige un número de usuarios de 250 o más.

4. Recorrido Cognitivo

El recorrido cognitivo (cognitive walkthrough) es un buen método para comenzar. Se trata de un método que no requiere usuarios, y es riguroso.

El ámbito del método es relativamente restringido. Se trata de investigar cómo interactuará un usuario con un producto en desarrollo en su primer contacto.

El método de evaluación supone que el usuario hará los siguientes pasos para lograr su tarea:

- 1) El usuario se pone un objetivo
- 2) El usuario busca las opciones disponibles en la interfaz de usuario
- 3) El usuario selecciona la acción que parece adecuada para progresar hacia el objetivo
- 4) El usuario realiza la acción y evalúa el progreso hacia los objetivos.

Para cada acción deben responderse las siguientes preguntas:

- a) ¿Tratará el usuario de lograr el efecto correcto? (selección correcta)
- b) ¿Notará el usuario que la acción está disponible? (visibilidad)
- c) ¿Asociará el usuario la acción correcta con el efecto a ser logrado? (rotulación)
- d) ¿Notará el usuario que se está progresando hacia el objetivo? (retroalimentación)

Se trata de un método analítico riguroso porque da instrucciones detalladas de cómo analizar los pasos del usuario hacia el resultado [2]. Lo puede aplicar casi cualquier persona, incluyendo a los desarrolladores.

Observe que este método puede ser utilizado con ventaja en una evaluación formativa. En particular, puede usarse en etapas tempranas de diseño, incluso con un prototipo de papel.

5. Inspección de usabilidad

La inspección de usabilidad es un método rápido que no utiliza usuarios. Recurre más bien a unos pocos “inspectores”: expertos en usabilidad independientes. Generalmente se aplica a prototipos.

Cada inspector, en forma individual, analiza el prototipo basado en su experiencia. Por supuesto, el resultado es dependiente de lo que el experto considera más relevante, pero fundamentado en principios de diseño o mejores prácticas. Es por eso, que las evaluaciones de dos inspectores no coinciden necesariamente. No obstante, se ha observado que, en muchos casos, el aporte dos o tres inspectores es suficiente para encontrar la mayoría de los problemas de un desarrollo.

Por supuesto, el método es cualitativo y formativo.

Note que si bien no necesitamos usuarios, requerimos el aporte de expertos en usabilidad independientes. Se les puede encontrar entre profesionales conocidos o a través de empresas que ofrezcan el servicio.

6. Evaluación heurística

Este método es similar a la inspección de usabilidad, pero se elimina la necesidad de recurrir a expertos de usabilidad. Es por ello que es considerado un método de descuento por excelencia. También por eso es muy popular.

En principio, en vez de expertos pueden actuar diseñadores, desarrolladores de software y/o expertos en el área de aplicación.

¿Cómo reemplazamos la experticia de los inspectores? Se pide a los evaluadores que se ciñan a una lista de 10 a 12 “heurísticas” o principios de diseño estándares. En particular, se pueden utilizar las heurísticas de usabilidad de Nielsen, que mencionamos en el capítulo 2.

Seis a ocho evaluadores examinan individualmente el prototipo, verificando si en él se verifican las heurísticas, y tomando todas las notas que estimen convenientes. En seguida, los evaluadores tienen una reunión en la que comparan sus notas y generan una lista combinada de problemas detectados. Note que la reunión debe ser participativa, y **no** debe ser la reunión en la que se apruebe el trabajo de un “encargado”.

El método es cualitativo, formativo, rápido y no requiere usuarios.

7. Focus groups

Los focus groups son una técnica bien conocida en las Ciencias Sociales. Consiste en organizar reuniones con 6 a 9 personas para conversar sobre un tema determinado. En el contexto de evaluación de usabilidad, se puede organizar un focus group tanto antes de que exista una interfaz como después de que haya estado usándose durante algún tiempo [3].

Desde el punto de vista del asistente, se trata de una reunión no estructurada, pero en realidad el moderador tiene un script preparado sobre qué preguntar.

Los focus groups pueden ser muy útiles para obtener reacciones e ideas espontáneas.

Con respecto a la preparación: se debe invitar no menos de 6 personas a un focus group, idealmente representativos de los usuarios futuros. También, es deseable tener dos grupos distintos (es decir, hacer dos focus groups en paralelo).

Hay que tener cuidado con las preguntas a hacer. Si se les pregunta “qué quieren”, se corre el riesgo de que los usuarios “piensen” que quieren algo cuando en realidad quieren otra cosa, pero responden basados en un escenario idealizado o simplemente inexistente. Es deseable, entonces, que los usuarios “experimenten” algo antes de participar en el focus group.

Durante la preparación de un focus group, el moderador debe hacer una lista de temas a discutirse, y establecer objetivos para el tipo de información a obtenerse.

Posteriormente, en la sesión, el moderador debe mantener foco de discusión, evitando que se mueva a otros temas.

También, el moderador debe tratar que todos los miembros contribuyan, y que nadie domine la discusión.

Finalmente, el análisis de datos es simple: el moderador escribe un informe resumiendo lo principal, incluyendo las citas más relevantes de lo dicho en la reunión.

El siguiente ejemplo muestra una aplicación de focus groups:

- Objetivo: se está diseñando un sistema avanzado de comunicaciones para trabajo en la oficina y el hogar.
- Primero: focus group de necesidades, con 6 usuarios discutiendo sus necesidades de comunicación con respecto a herramientas y prácticas actuales.
- En seguida, un grupo de expertos diseñan 5 opciones de diseño y escenarios de uso para satisfacer las necesidades identificadas.
- Finalmente, se cita de nuevo al focus group de 6 usuarios para discutir las propuestas del sistema futuro.

Los focus groups son un método rápido, cualitativo, formativo, y requiere usuarios.

8. Mago de Oz

Este método se inspira en la magnífica película “El mago de Oz” (Wizard of Oz) de 1939 (<https://youtu.be/YWyCCJ6B2WE>), considerada por la UNESCO como Memoria del Mundo.

Como se sabe, en la película el supuesto mago no es tal, sino que esta persona manipula “por detrás” lo que el público ve. Esa idea es la que se toma en este método, que se aplica a desarrollos en los que aún faltan componentes. Por ejemplo, piense en un desarrollo en que la interfaz de usuario envía un requerimiento a un servidor para que provea una cierta información. Si el servidor aún no está operativo, se puede tener un “mago” que provea esa cierta información. Por supuesto, hay algo de engaño en el hecho de que esa información no está siendo realmente obtenida por computador, sino que alguien pone alguna información estándar o previamente determinada “como si” la proveyera el servidor. Por su parte, el usuario frente a la interfaz no tiene por qué saber esto si el mago está suficientemente oculto; su interés es ver si todo es accesible, fácil de usar, se presenta bien, etc., a través de la interfaz.

En consecuencia, se trata de un método rápido, con usuarios, apropiado para etapas intermedias de diseño, en las cuales ciertas funcionalidades no son fáciles de prototipar, y que son asumidas por un ser humano. La manera típica de funcionamiento es que el sistema en evaluación funciona en un computador conectado a otro, donde una persona hace las funciones emuladas. El emulador no es visible al usuario, quien cree que el sistema es funcional, lo que ayuda a hacer una evaluación más naturalista. Note que se necesita desarrollar una interfaz de usuario (lo más realista posible) y otra interfaz para el emulador.

Ejemplo: suponga que se está desarrollando una interfaz para conductor de vehículo, que presenta al usuario la velocidad máxima en la ruta que está conduciendo. La siguiente figura muestra un usuario que está evaluando dicha interfaz, con un emulador que sobrepone en la proyección un slide con el símbolo de la velocidad máxima (en este caso, el mago está muy visible).



9. Pensar en voz alta

“Thinking aloud” es un método clásico de evaluación de usabilidad. Es con usuarios, de descuento y riguroso. A cada usuario se le pide que utilice el software para resolver un problema mientras va diciendo en voz alta lo que está pensando y haciendo. El evaluador observa directamente y toma notas sobre los dichos y acciones del usuario.

Este método provee un acceso privilegiado a cómo los usuarios utilizan una interfaz. Por lo tanto, es muy útil para evaluar fenómenos cognitivos complejos, tales como percepción de situación (situation awareness), toma de decisiones, y uso de memoria.

El mejor momento para ejecutar una prueba es al principio del proceso de diseño, antes de avanzar demasiado. Esto podría ocurrir cuando tenga un prototipo sólido o cuando una parte breve pero representativa de su producto esté completa.

Hay dos tipos comunes de pruebas de usabilidad de pensar en voz alta: verbalización concurrente y retrospectiva. Los participantes completan una tarea y narran lo que pasa por su mente. En el método de pensar en voz alta retrospectivo, los participantes informan sobre su experiencia después de completar la tarea.

Los pasos para ejecutar una prueba de usabilidad de Think Aloud son:

1. Metas. Identificar los objetivos de la prueba de usabilidad.
2. Reclutamiento. Reclute diversos participantes de su público objetivo. Idealmente, tendrán diversos roles laborales, antecedentes, educación, etc.
3. Tareas. Diseñe tareas realistas para que las complete el participante. En productos de aprendizaje, estas podrían ser actividades como pedir a los participantes que:
 - Inicie sesión en el LMS, busque el curso XYZ e inicie el curso.
 - Repase la Lección 2 y complete las actividades.
 - Complete la autoevaluación y envíese los resultados por correo electrónico.
 - Encuentre el recurso sobre diseño gráfico y descárguelo.
4. Guion. Escriba un guion para leer al comienzo de la prueba. Comience con una introducción amistosa para ayudar al participante a sentirse cómodo. Es posible que desee comenzar con una afirmación como: "Estamos probando el producto, no usted". Preséntese y pregunte sobre el trabajo o función de la persona. Explique cómo funcionará la prueba.
5. Corre. Realice la prueba en persona o mediante una herramienta de videoconferencia. Esté preparado para pedirle al participante que solo informe lo que está pensando.
6. Perspectivas. Reúna sus datos en un solo lugar y busque patrones.
7. Presente. Presente sus hallazgos para ayudar a que todos tengan un entendimiento común. Por lo general, este es su equipo y las partes interesadas.

8. Revisar. Revise su producto de aprendizaje para aliviar los problemas que identificaron los participantes.

10. Introducción a métodos basados en cuestionarios

Los cuestionarios son ampliamente utilizados en las Ciencias Sociales. Es por ello que se han desarrollado también para usarse en la evaluación de usabilidad de software. Una primera observación que es necesario hacer es que los cuestionarios necesitan la participación de usuarios. La segunda es que los cuestionarios deben ser respondidos por los usuarios después de usar el software en cuestión. Esto parece obvio, sin embargo, hemos visto memorias en que se ha aplicado la encuesta después que los usuarios han visto un video de uso, lo cual es incorrecto. También, deseablemente, deben responder el cuestionario inmediatamente después de haber usado el software.

Los usuarios a los que se aplica los cuestionarios deben responder todas las preguntas. Normalmente, las preguntas tienen respuestas codificadas (por ejemplo, “de acuerdo”, “no lo sé”, “en desacuerdo”), por lo que responder las preguntas es muy rápido, incluso en aquellos cuestionarios con más preguntas.

Cada cuestionario necesita un número mínimo de usuarios. Como recomendación, es importante considerar inicialmente más usuarios que el mínimo, debido a que muchas veces un usuario no responde (aún después de asegurar que lo haría) o responde muy pocas de las preguntas.

Siguiendo las mejores prácticas, el uso de estos cuestionarios puede generar evaluaciones rigurosas. En caso contrario, resultarán sólo evaluaciones rápidas.

Hay cuestionarios con 50 preguntas (como SUMI), otros con 30 (como USE), etc. ¿Será posible conseguir un cuestionario riguroso con menos preguntas? Esta pregunta nos recuerda la negociación de Abraham con Dios [4]: inicialmente la ciudad de Sodoma se salvará si hay 50 justos en la ciudad. Abraham pregunta si con 45 justos podrá Dios perdonar a la ciudad. Y así continúa, reduciendo el número de justos necesario, hasta negociar sólo 10 justos. Lamentablemente, en Sodoma no existe ni siquiera ese número de justos y la ciudad es destruida. ¿Se podrá crear un cuestionario riguroso con sólo 10 preguntas? Veámoslo en los siguientes capítulos. Finalmente, es necesario mencionar que la utilización de algunos cuestionarios es pagada. Por ejemplo, disponer del cuestionario SUMI y su metodología de evaluación, que hemos mencionado, es con costo en dinero.

11. Cuestionario USE

El cuestionario USE (Usefulness, Satisfaction and Ease of use) fue propuesto por Lund en 2001 [5]. Consiste de 30 preguntas, que evalúan cuatro dimensiones de usabilidad: utilidad (UU), facilidad de uso (UE), facilidad de aprendizaje (UL), y satisfacción (US). Cada pregunta se responde en una escala Likert de 7 valores (1 = “strongly disagree”, a 7 = “strongly agree”). A continuación se muestran las 30 preguntas y su respectiva dimensión.

- UU1 It helps me be more effective.
- UU2 It helps me be more productive.
- UU3 It is useful.
- UU4 It gives me more control over the activities in my life.
- UU5 It makes the things I want to accomplish easier to get done.
- UU6 It saves me time when I use it.
- UU7 It meets my needs.
- UU8 It does everything I would expect it to do.
- UE1 It is easy to use.
- UE2 It is simple to use.
- UE3 It is user friendly.
- UE4 It requires the fewest steps possible to accomplish what I want to do with it.
- UE5 It is flexible.
- UE6 Using it is effortless.
- UE7 I can use it without written instructions.
- UE8 I don't notice any inconsistencies as I use it.
- UE9 Both occasional and regular users would like it.
- UE10 I can recover from mistakes quickly and easily.
- UE11 I can use it successfully every time.
- UL1 I learned to use it quickly.
- UL2 I easily remember how to use it.
- UL3 It is easy to learn to use it.
- UL4 I quickly became skillful with it.
- US1 I am satisfied with it.
- US2 I would recommend it to a friend.
- US3 It is fun to use.
- US4 It works the way I want it to work.
- US5 It is wonderful.
- US6 I feel I need to have it.
- US7 It is pleasant to use.

¿Cómo interpretamos los resultados? Podemos promediar los resultados obtenidos en cada dimensión y averiguar cómo le va a nuestro producto. Obviamente, un producto que tenga promedio sobre 5 en las respuestas a las preguntas es un buen producto. Si se obtiene bajo 4 en alguna dimensión significa evaluación negativa en esa categoría.

El método USE es considerado riguroso con un número apropiado de usuarios. Nosotros recomendamos 20 o más.

Entre los aspectos positivos del método USE encontramos que es un método que permite distinguir, en la usabilidad, las cuatro dimensiones. Por ejemplo, se puede evaluar claramente el tema facilidad de aprendizaje. También, que tiene alta confiabilidad debido en parte a que tiene preguntas que de alguna forma son redundantes. Por otra parte, como tiene muchas preguntas, es relativamente independiente de las palabras exactas que se usen para las preguntas, lo cual es útil para traducir con seguridad el cuestionario a otro idioma. Finalmente, como tiene la misma escala numérica que nuestro sistema de notas en Chile (1 a 7), es fácil para los usuarios responder el cuestionario y para los evaluadores apreciar el resultado. Entre los aspectos negativos de USE está el hecho de que todas las preguntas están formuladas en un sentido “positivo”, lo cual induce de alguna manera las respuestas de los usuarios. También, se ha formulado alguna duda respecto a la invariancia del cuestionario a los grupos a los que se aplica [6]. Esta última es la razón por la que sugerimos un número mínimo de usuarios relativamente alta.

12. Cuestionario PSSUQ

El cuestionario PSSUQ (Post Study Usability Questionnaire) es un instrumento de 16 preguntas. Utiliza una escala Likert de 7 valores, pero a diferencia de USE, 1 es “Strongly agree”, y 7 es “Strongly disagree”. Por lo tanto, no se puede hacer equivalencia con la escala de notas que usamos en Chile.

PSSUQ fue desarrollado en la década de 1980 por Lewis [7].

The Post-Study Usability Questionnaire Version 3			Strongly agree							Strongly disagree	NA
			1	2	3	4	5	6	7		
1	Overall, I am satisfied with how easy it is to use this system.		☐	☐	☐	☐	☐	☐	☐		☐
2	It was simple to use this system.		☐	☐	☐	☐	☐	☐	☐		☐
3	I was able to complete the tasks and scenarios quickly using this system.		☐	☐	☐	☐	☐	☐	☐		☐
4	I felt comfortable using this system.		☐	☐	☐	☐	☐	☐	☐		☐
5	It was easy to learn to use this system.		☐	☐	☐	☐	☐	☐	☐		☐
6	I believe I could become productive quickly using this system.		☐	☐	☐	☐	☐	☐	☐		☐
7	The system gave error messages that clearly told me how to fix problems.		☐	☐	☐	☐	☐	☐	☐		☐
8	Whenever I made a mistake using the system, I could recover easily and quickly.		☐	☐	☐	☐	☐	☐	☐		☐
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.		☐	☐	☐	☐	☐	☐	☐		☐
10	It was easy to find the information I needed.		☐	☐	☐	☐	☐	☐	☐		☐
11	The information was effective in helping me complete the tasks and scenarios.		☐	☐	☐	☐	☐	☐	☐		☐
12	The organization of information on the system screens was clear.		☐	☐	☐	☐	☐	☐	☐		☐
13	The interface* of this system was pleasant.		☐	☐	☐	☐	☐	☐	☐		☐
14	I liked using the interface of this system.		☐	☐	☐	☐	☐	☐	☐		☐
15	This system has all the functions and capabilities I expect it to have.		☐	☐	☐	☐	☐	☐	☐		☐
16	Overall, I am satisfied with this system.		☐	☐	☐	☐	☐	☐	☐		☐

*The “interface” includes those items that you use to interact with the system. For example, some components of the

PSSUQ tiene tres dimensiones de usabilidad: Calidad del sistema (promedio de las respuestas a las preguntas 1 a la 6), Calidad de la información (promedio de las respuestas a las preguntas 7 a la 12), y Calidad de la interfaz (promedio de las respuestas a las preguntas 13 a 16). Por lo tanto, se puede obtener evaluaciones de cada una de estas dimensiones, y el promedio total para la evaluación general. Ahora, mientras más baja la puntuación, mejor es la evaluación. Note que un puntaje 4 no necesariamente significa “usabilidad promedio”, sino simplemente que

los usuarios consideran que la usabilidad del sistema está equidistante entre “alta” y “baja”.

El cuestionario es apropiado para ser usado tanto con muestras pequeñas como grandes (más de 100 usuarios). La única diferencia es el grado de precisión que se obtiene. Tullis y Stetson [8] estudiaron con el método la usabilidad de dos sitios web financieros; encontraron que una muestra de sólo 12 usuarios generó los mismos resultados que una muestra mucho mayor 90% de las veces. Nosotros recomendamos al menos 15 usuarios para estudios sumativos en el caso de una muestra pequeña.

El cuestionario PSSUQ es altamente confiable, y es un método riguroso.

Otro aspecto positivo adicional de este cuestionario es que provee tres dimensiones de evaluación, además de la evaluación general.

Con respecto a los aspectos negativos, tenemos los siguientes. Las preguntas están todas presentadas en forma positiva, lo que induce de alguna manera al usuario en esa dirección. El orden de presentación de las respuestas (de máxima coincidencia a mínima coincidencia) es contraintuitivo; de hecho, el mismo Lewis ha dicho que si él pudiera cambiar algo al método, sería poner el orden contrario.

13. El cuestionario SUS

El cuestionario SUS (System Usability Scale) fue introducido en 1996 por Brooke [9]. Aunque lo presentó como un método “quick and dirty” (rápido y sucio), SUS puede ser riguroso si se siguen las mejores prácticas.

Utiliza sólo diez preguntas, que se responden en una escala Likert de 5 valores. Esta economía de recursos ha hecho que el método sea muy popular, aunque muchas veces se aplica incorrectamente.

Las preguntas son las siguientes:

The System Usability Scale Standard Version		Strongly disagree					Strongly agree				
		1	2	3	4	5					
1	I think that I would like to use this system.		☐	☐	☐	☐	☐				
2	I found the system unnecessarily complex.		☐	☐	☐	☐	☐				
3	I thought the system was easy to use.		☐	☐	☐	☐	☐				
4	I think that I would need the support of a technical person to be able to use this system.		☐	☐	☐	☐	☐				
5	I found the various functions in the system were well integrated.		☐	☐	☐	☐	☐				
6	I thought there was too much inconsistency in this system.		☐	☐	☐	☐	☐				
7	I would imagine that most people would learn to use this system very quickly.		☐	☐	☐	☐	☐				
8	I found the system very cumbersome to use.		☐	☐	☐	☐	☐				
9	I felt very confident using the system.		☐	☐	☐	☐	☐				
10	I needed to learn a lot of things before I could get going with this system.		☐	☐	☐	☐	☐				

Como puede observarse, las preguntas se alternan en ser formuladas positivamente y negativamente. Las respuestas se codifican en valores de 0 a 4 (es decir, valor de formulario menos 1). Para calcular el puntaje que otorga un usuario a su cuestionario, se realizan los siguientes cálculos:

- Se suman los puntajes de las respuestas a las preguntas impares.
- Para las preguntas pares, se suman los valores $(4 - \text{puntaje})$
- Se suman los sub-totales anteriores. Esto da un valor entre 0 y 40.
- Se multiplica este valor por 2.5. Como resultado, se obtiene un valor entre 0 y 100.

En seguida, se calcula el promedio de los puntajes individuales así obtenidos. Nuevamente, se obtiene un valor entre 0 y 100. Esto sería un indicador de la usabilidad del sistema, pero no debe interpretarse linealmente. En cambio, debe

compararse con un valor “límite”, que corresponde al promedio de un gran número de evaluaciones realizadas en EE.UU. Si el valor calculado es mayor al límite, se considera que el sistema tiene una usabilidad positiva; si no, tiene una usabilidad negativa. El límite es el valor 68.

En consecuencia, SUS no entrega en realidad un valor numérico, sino una respuesta de tipo “sí” o “no”. En muchos casos, esto es suficiente: saber si el sistema tiene una buena usabilidad (o no). Ahora, si se necesita saber más dimensiones de usabilidad, hay que recurrir a otro cuestionario. No es correcto comenzar a analizar las respuestas de los usuarios a alguna de las diez preguntas individuales, especialmente cuando la muestra de usuarios es pequeña. No obstante lo anterior, recientemente se ha insinuado que se podría usar las respuestas a las preguntas 4 y 10 para tener una aproximación a la dimensión “facilidad de aprendizaje”.

Con respecto a tamaño de la muestra, la literatura menciona un tamaño mínimo de 5 a 15 usuarios. Los 5 usuarios son propuestos por Brooke mismo, “cuando las respuestas de los usuarios son consistentemente similares”. Nosotros preferimos poner un límite mínimo de 10 usuarios y así, no preocuparse de qué significa exactamente “respuestas similares”.

Este tamaño mínimo de la muestra de sólo 10 usuarios atrae a muchos evaluadores que tienen problemas para encontrar suficientes usuarios.

Por otro lado, debido a su propuesta minimalista, SUS adolece de una debilidad: la redacción de las diez preguntas tiene impacto en los resultados que se obtienen. Se ha demostrado en EE.UU. que cambiando palabras de las preguntas por sinónimos produce resultados (puntajes) distintos a los originales. Esto tiene implicancias en el caso de utilizar el cuestionario en otros países: distintas traducciones de las preguntas producirán resultados diferentes, y el límite de 68 tampoco, en principio, sería válido. En varios países se ha estandarizado una cierta traducción y por lo tanto, se conoce el nuevo límite (que puede coincidir con el número 68), y en consecuencia, se puede usar con entera confianza.

Lamentablemente, en Chile, a nuestro conocimiento, no existe una traducción estándar con las características mencionadas previamente. ¿Significa entonces que en Chile no se puede traducir el cuestionario y hay que usar la versión original en inglés? No, no sería apropiado, ya que el nivel promedio de conocimiento de inglés en el país es muy bajo (más bajo incluso que muchos países de América Latina), y en ningún caso similar el que se tiene en EE.UU.

¿Se puede hacer algo, entonces? Una solución parcial es la siguiente:

Se puede hacer una traducción ad-hoc del cuestionario. En seguida, aplicarlo. Si el puntaje resultante es alto, por ejemplo, 88, entonces se puede afirmar con

propiedad que el resultado de la usabilidad es positivo. El efecto de la traducción no podría ser responsable de 20 puntos de diferencia entre puntaje calculado y el límite.

Un problema ocurre si la diferencia entre puntaje calculado y límite es sólo de 4 o 5 puntos. En este caso, sí podría haber un efecto significativo de la traducción. En esta situación, hay que prescindir de la pretensión de que la evaluación es rigurosa y calificarla como evaluación rápida. Si se necesita una evaluación rigurosa, entonces hay que hacer una evaluación con otro método.

Resumiendo, SUS puede ser usado como un método riguroso de evaluación a partir de 10 usuarios y con sólo 10 preguntas. Sin embargo, tal como en la huida de Lot y su familia desde Sodoma antes de su destrucción, hay un problema: no se puede mirar hacia atrás (o en el caso chileno de utilización de SUS, hay dificultad de interpretación cuando el puntaje promedio calculado de las respuestas es cercano al límite).

14. Testeo de usabilidad de guerrilla

Este método rápido es útil para obtener de manera barata retroalimentación de usuarios. Se hace con usuarios cualesquiera: en la calle, una cafetería, o una biblioteca, etc. Para convencer en la invitación a usuarios, se puede dar un pequeño incentivo: una taza de té o un pastel, por ejemplo. La respuesta se espera rápida, por lo que se deben preguntar muy pocas cosas.

Esto es apropiado para obtener información de usuarios del tipo que habrá en el lugar: si es en la calle, será usuarios con características indeterminadas (salvo género y rango etario); si es en una cafetería universitaria, será de estudiantes universitarios. Normalmente, se pide reacciones a un número de personas entre 8 y 12.

En general, se usa cuando se va a proponer una interfaz muy novedosa y el diseñador no tiene idea de cómo reaccionarán los usuarios, y las barreras que podrían existir. Como es muy barato de aplicar, el evaluador puede usar el método las veces que desee. Por otra parte, note que este método es de grano grueso: quizás no proporcione los detalles de usabilidad que puedan necesitarse para afinar un diseño.

15. Aceptación de tecnología: modelo TAM

Suponga que una empresa o institución va a cambiar un software importante que usan los empleados. ¿Cómo reaccionarán los empleados? Note que, de partida, es normal que haya una reacción inicial negativa: cada persona deberá desechar el conocimiento del actual sistema para empezar a aprender un nuevo sistema. Ahora, eso es una etapa de transición, que ojalá sea corto, y después, dependerá de cómo los empleados valoran el nuevo software en comparación con el antiguo...

Siempre existe el riesgo de que la valoración sea inferior. ¿Podemos predecir o anticipar cuál será la aceptación de la nueva tecnología? Eso es lo que pretende Davis, proponente del modelo TAM (Technology Acceptance Model) [10], basado en teorías psicológicas.

TAM utiliza las relaciones entre las siguientes variables:

Perceived Usefulness (PU): utilidad percibida. Se define como el grado con el que una persona cree que usando el sistema objetivo mejorará su desempeño.

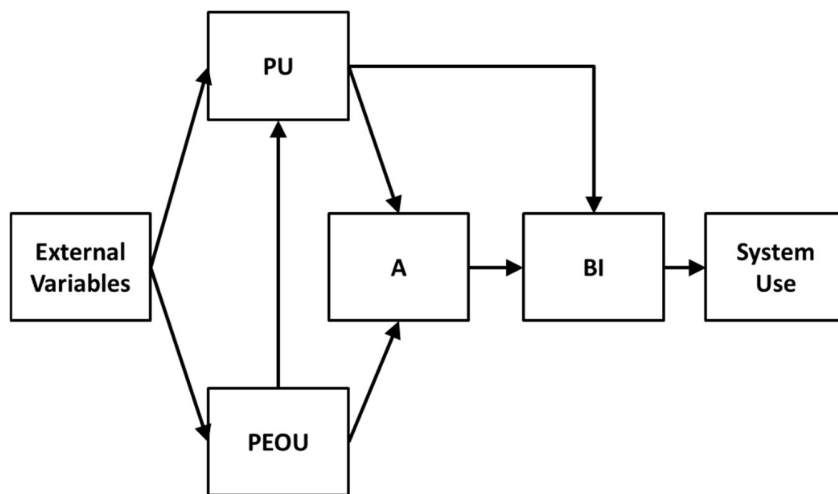
Perceived Ease-of-Use (PEOU): la facilidad de uso percibida. Se define como el grado en que una persona cree que usando el sistema objetivo estaría libre de esfuerzo.

Attitude Toward Behavior (A): actitud hacia el comportamiento. Los sentimientos (positivos o negativos) acerca de realizar el comportamiento objetivo.

Behavioral Intention (BI): intención de comportamiento. El grado en que una persona ha formulado planes conscientes para realizar o no algún comportamiento futuro especificado.

Estas serían las variables principales relacionadas con la aceptación de tecnología. Hay otras variables, conocidas como variables externas, que afectan el uso del sistema. Ejemplos de variables externas: diseño del sistema, atributos del usuario, características de la tarea, estructura organizacional. ¡Se han identificado 39 factores que afectan la satisfacción de sistemas de información! [11].

TAM propone que el pronóstico de uso del sistema se obtiene de acuerdo a lo mostrado en la siguiente figura:



TAM establece que las variables externas inciden sobre PEOU. PU es afectado por variables externas y PEOU. A es afectado por PU y PEOU. El uso del sistema es determinado por BI, que es afectado por A y PU.

Entonces, si es posible determinar o estimar las variables, es posible estimar el uso del sistema.

De hecho, existen usos exitosos de TAM publicados. Pero también, el modelo ha sido criticado porque ignora muchos aspectos tales como el relacionamiento social entre los usuarios, los costos, los imperativos que fuerzan a los usuarios.

En todo caso, el impacto de TAM en el área de Sistemas de Información fue enorme. La idea de que se pueda predecir la aceptabilidad de una nueva tecnología es muy atractiva.

Posteriormente a TAM, se desarrollaron una variedad de otros modelos, que de alguna forma corregían las deficiencias de TAM. Por ejemplo, TAM2, TAM3, TTM, UTAUT, EBTAM.

UTAUT [12] es considerado hoy como el modelo más completo. EBTAM lo explicamos en el siguiente capítulo.

16. Aceptación de tecnología: modelo EBTAM

En el capítulo anterior, mencionábamos que UTAUT es considerado un modelo muy completo. Sin embargo, su aplicación es apropiada en el caso de grandes empresas. Si se trata de pequeñas o medianas empresas, UTAUT es demasiado oneroso.

EBTAM (E-Business Technology Acceptance Model), propuesto por Leyton y otros en 2015 [13] también tiene su inspiración en TAM, pero está orientado a las empresas más pequeñas, con menos recursos a destinar al estudio de evaluación.

Se propone, como en TAM, la influencia de PU y PEOU sobre BI. Se evita el uso de A. Se evita la utilización de USE como medida de éxito, cuando el uso de la tecnología es una obligación (como es el caso habitual en empresas, a diferencia, por ejemplo, de un estudio de profesionales, donde el uso de cierta tecnología puede ser optativo). Estudiemos los temas de evaluar PU y PEOU en este modelo.

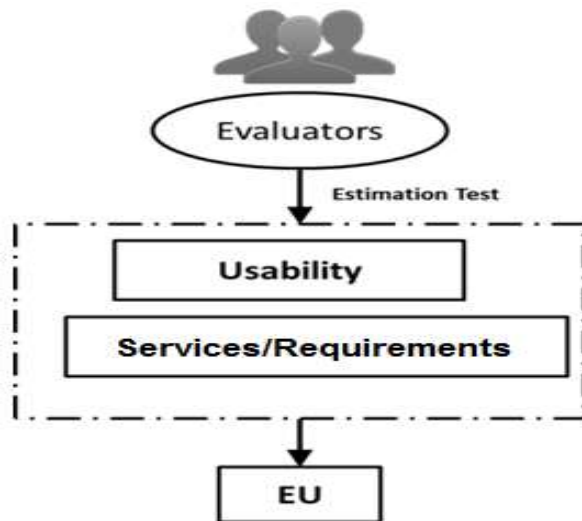
La PU es difícil de calcular cuando los usuarios no tienen experiencia previa con el sistema o tecnología. Esto ocurre, por ejemplo, cuando los usuarios pasan de negocios en papel a negocios vía electrónica. Es mejor que unos pocos expertos determinen el valor de la tecnología (porque la conocen). Estos **evaluadores** trabajan de manera similar al proceso Wideband Delphi para estimar procesos de software (tienen que estar de acuerdo si los requerimientos de la solución se proveen a través de los servicios ofrecidos por la herramienta bajo evaluación).

La PU se llama ahora **Estimated Utility (EU)**.

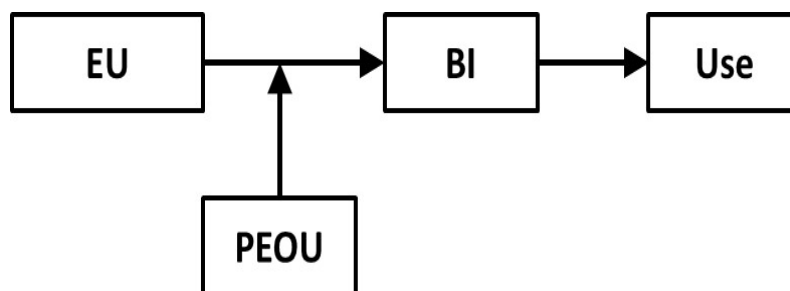
Para calcular la EU se puede usar la matriz de trazabilidad sobre los requerimientos funcionales. Se puede también ponderar los requerimientos, para tomar en cuenta que hay algunos más importantes que otros. Una escala 0 a 4 es recomendada.

También se debe estudiar la **Usabilidad Esperada (EUSA)**. Esta puede ser 1 (cumple mínimos) o 0 (no cumple).

En resumen, la **EU** puede calcularse como muestra la siguiente figura.



Veamos ahora cómo calcular la Facilidad de Uso Percibida (PEOU). Recordemos el uso de PEOU:



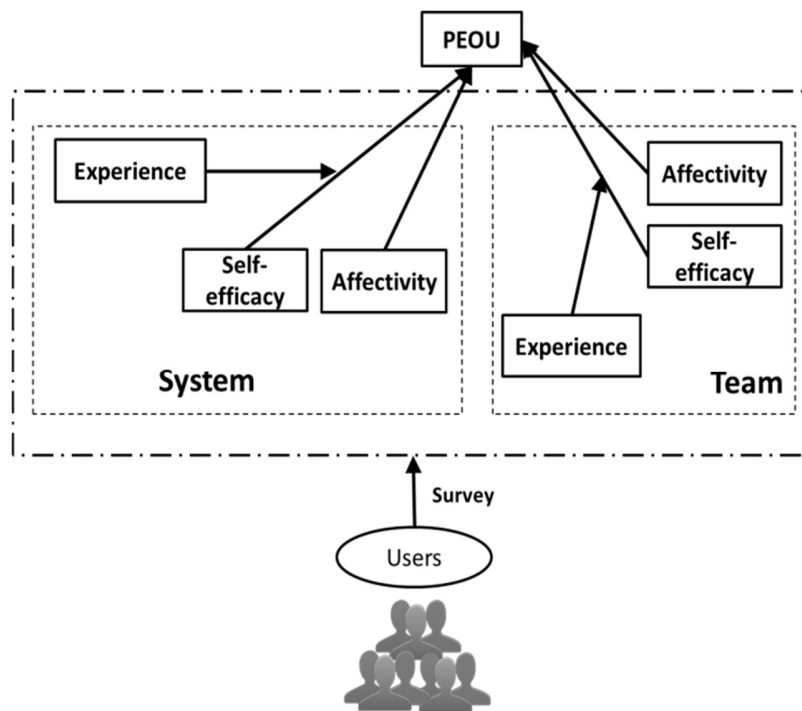
La Facilidad de Uso Percibida es considerada en EBTAM un factor secundario cuando se la compara con EU. Esto se puede ver a través de un ejemplo. Supongamos dos sistemas: A y B. Si A cumple los requerimientos, pero B no, no vale la pena evaluar la PEOU de B.

De TAM 3, el modelo EBTAM rescata las variables *anxiety* y *self-efficacy*.

Anxiety: tiene relación con la afectividad, y así, se la reemplaza por **Positive affectivity**.

Self-efficacy tiene mucho que ver con experiencia con la tecnología en análisis. La variable **Experience** queda moderando.

En resumen, tenemos para PEOU:



En la figura anterior, aparecen dos bloques: “Sistema” y “Equipo”, pero en realidad, después de evaluaciones, se concluye que la componente “Equipo” no es relevante, por lo que el modelo se queda al final con “Sistema” solamente.

En la propuesta original, el modelo se evaluó en tres casos, mostrando utilidad [13]. Un aspecto a tomar en cuenta es que EBTAM es menos preciso que modelos más completos como UTAUT. Sin embargo, es muchísimo más liviano. Esta es la razón por la cual es útil para evaluaciones rápidas en pequeñas y medianas empresas, como se ha comprobado en varios casos.

Bibliografía

1. Nielsen, J.: "Heuristic evaluation". In J. Nielsen and R.L. Mack (eds.): Usability Inspection Methods. J. Wiley & Sons, New York, 1994.
2. Rieman, J., Franzke, M. and Redmiles, D.: "Usability evaluation with the cognitive walkthrough". In Conference companion on Human factors in computing systems (pp. 387-388). ACM, 1995.
3. Caplan, S.: "Using focus groups methodology for ergonomic design". Ergonomics 33, 5, 527-533, 1990.
4. Sagrada Biblia. Génesis 18:16-33.
5. Lund, A. M.: "Measuring Usability with the USE Questionnaire". Usability Interface, 8(2), 3-6, 2001.
6. Gao, M., Kortum, P., and Oswald, F.: "Psychometric Evaluation of the USE (Usefulness, Satisfaction and Ease of Use) Questionnaire for Reliability and Validity". Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting 1414-1418.
7. Lewis, J.D.: Using the PSSUQ and CSUQ in User Experience Research and Practice. Measuring U, 2019.
8. Tullis, T. and Stetson, J. "A Comparison of Questionnaires for Assessing Website Usability", 2006.
9. Brooke, J.: "SUS-A quick and dirty usability scale". Usability evaluation in industry, 189(194), pp.4-7, 1996.
10. Davis, F.D. "Perceived usefulness, perceived ease of use, and user acceptance of information technology". MIS Quarterly 13 (3), 319-340, 1989.
11. Bailey, J.E. and Pearson, S.W.: "Development of a Tool for Measuring and Analyzing Computer User Satisfaction". Management Science 29, 530-545, 1983.
12. Venkatesh, V., Morris, M.G., Davis G.B., and Davis, F.D.: "User acceptance of information technology: Toward a unified view". MIS Quarterly 27(3), 425-478, 2003.
13. Leyton, D., Pino, J.A., and Ochoa, S.F.: "EBTAM: Technology Acceptance in e-Business Environments", Information Systems and e-Business Management 13 (2), 211-234, 2015.

Agradecimientos

Los autores agradecen al Prof. Pedro Antunes la inspiración y ejemplos en la escritura de esta Guía.