



TEMA 7

Evaluación

Diseño Centrado en el Humano y Experiencia de Usuario
Curso 2024-2025

Máster Universitario en Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid

Contenido

7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

7.2 Técnicas de Evaluación de la Usabilidad

7.3 Experimentación con Usuarios

Bibliografía y Referencias Específicas

7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

- **Evaluación**

- Parte integrante del modelo de proceso de ingeniería de usabilidad y accesibilidad
- Actividad necesaria de comprobación, formada por un conjunto de metodologías y técnicas que analizan la usabilidad y la accesibilidad de un sistema interactivo en **diferentes etapas** del ciclo de vida del software

- **Lugares** de realización

- En entorno de trabajo
- En laboratorios especializados

- **Objetivos** de la evaluación

- Comprobar aspectos como: facilidad de aprendizaje, manejabilidad, tiempo consumido, ...
- Identificar problemas específicos con el sistema (resultados inesperados, ...)

- ¿**Cuánto** cuesta una evaluación y qué obtendremos con su realización?

7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

- **Laboratorio de usabilidad**

- Normalmente
 - Sala de observación (para evaluadores) y sala de pruebas (para usuarios), separadas por un cristal que permite ver en un solo sentido
- Permanentes/móviles
- Opcionalmente
 - Cámaras de vídeo
 - Rastreo ocular (*eye tracking*)

- **Laboratorio de accesibilidad**

- Más complicado de montar



7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

- Clasificación de métodos de evaluación según **criterio “lugar de evaluación”**
 - **1) Laboratorio**
 - Ventajas
 - Sofisticados equipos. Los usuarios pueden operar sin distracciones.
 - Inconvenientes
 - Las distracciones forman parte del entorno natural de trabajo y pueden tener una influencia significativa. No se reproduce el contexto de la interacción.
 - **2) Entorno de trabajo**
 - Ventajas
 - Es la única manera de ver realmente cómo los usuarios realizan sus tareas cooperando entre ellos. Mejora el análisis de requisitos (surgen aspectos nuevos).
 - Inconvenientes
 - El lugar de trabajo puede dificultar la observación (entorno más ruidoso, se puede alargar el tiempo de duración de la prueba, ...)

7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

- Clasificación de métodos de evaluación según **criterio “técnica de comprobación”**
 - **1) Métodos de inspección**
 - Los evaluadores (denominados inspectores) pueden ser especialistas en usabilidad, consultores de desarrollo de software, etc. El resultado final de evaluación se basa en el la opinión de los inspectores. Son generalmente métodos más baratos pues **no intervienen usuarios finales**.
 - **2) Métodos de indagación**
 - Los evaluadores interaccionan con los usuarios finales y los observan en su trabajo real. Se generan nuevas ideas de diseño. La **interacción con usuarios finales** se realiza a través de respuestas a preguntas, verbalmente o por escrito.
 - **3) Métodos de test**
 - Los evaluadores realizan **pruebas/tests con usuarios finales**, quienes utilizan el sistema (o prototipo)

7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

- Clasificación de métodos de evaluación según **criterio “automatización”**
 - **1) Automáticos.** El inconveniente es que se basan en sistemas ya finalizados (o versiones preliminares de lanzamiento) y no pueden usarse en etapas tempranas del ciclo de vida del desarrollo.
 - **2) Manuales.** El inconveniente es que se consume más tiempo y se necesitan más evaluadores. La ventaja es que son realizables en cualquier etapa del desarrollo y con cualquier tipo de prototipo.
- Clasificación de métodos de evaluación según **criterio “tipo de participantes”**
 - **1) Con usuarios e implicados**
 - Puede no ser fácil el acceso a usuarios e implicados del sistema
 - **2) Sin usuarios ni implicados**
 - Realizada por evaluadores con la ayuda de guiones, pautas, etc. Son más fáciles de implementar y detectan errores funcionales o de consistencia, pero carecen de la valiosa información de los usuarios.

7.1 Actividades Específicas en la Evaluación de Sistemas Interactivos

- Los **modelos de proceso** ([ejemplo](#)) plantean recomendaciones de desarrollo para sistemas interactivos, y en particular establecen **pautas de evaluación**:
 - **Acotar** claramente la parte del sistema a evaluar dentro del sistema global
 - Cada evaluación debe tener un **responsable** (miembro del equipo de desarrollo), cuyas funciones serán:
 - Moderar sesiones, reclutar usuarios y evaluadores, reservar lugar de evaluación, dirigir las tareas que se deben evaluar, disponer de todo lo necesario (prototipos, dispositivos, etc.), concretar las tareas a evaluar y las técnicas de evaluación que se utilizarán, etc.
 - La evaluación debe encajar en la **planificación y el presupuesto** del proyecto
 - Se deben **documentar** los resultados de evaluación y distribuirlos entre los miembros de equipo de desarrollo

7.2 Técnicas de Evaluación de la Usabilidad

- Métodos de evaluación de la usabilidad según criterio “técnica de comprobación”
 - **INSPECCIÓN. Técnicas:**
 1. Evaluación heurística
 2. Recorrido cognitivo (*cognitive walkthrough*)
 3. Inspección de estándares
 - Indagación
 - Test



Sin intervención de usuarios finales

7.2 Técnicas de Evaluación de la Usabilidad. Evaluación heurística

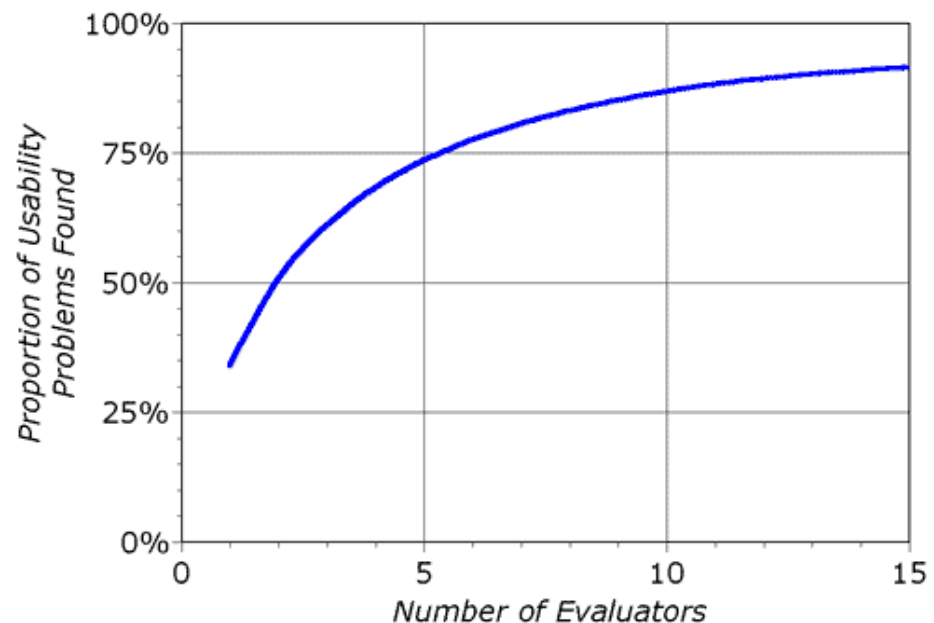
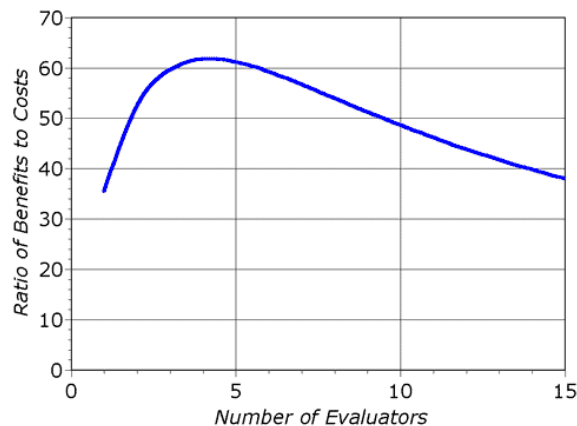
- Es el método de evaluación **más usado** (Molich y Nielsen, 1990)
- Está basado en la conformidad de la interfaz con ciertos principios reconocidos de usabilidad (**heurísticos**)
- Características
 - **Bajo coste**
 - Usable también en **etapas iniciales** del ciclo de vida
 - **Intuitivo y fácil de motivar** a evaluadores
- La inspección se realiza por **varios evaluadores**
 - El número recomendable es 3-5
 - Un evaluador puede ser un experto en usabilidad o un desarrollador

7.2 Técnicas de Evaluación de la Usabilidad. Evaluación heurística

- Cada evaluador actúa por separado, comunicando los resultados a un **observador común**
 - El responsable de la evaluación lee los informes de los evaluadores y diseña un documento de integración sobre la evaluación con recomendaciones para diseñadores
- Las heurísticas más usadas son las **10 reglas heurísticas de usabilidad de Nielsen** ([URL](#))
 - Conjunto revisado a partir del análisis de 249 problemas de usabilidad

7.2 Técnicas de Evaluación de la Usabilidad. Evaluación heurística

- Referencias: (Nielsen [URL1](#), AIPO [URL2](#) cap 15)
- El **número de evaluadores** expertos recomendable es 3-5 (Nielsen y Landauer, 1993), aunque para menos expertos la recomendación estaría en 10-12, [URL](#)



7.2 Técnicas de Evaluación de la Usabilidad. Evaluación heurística

- **Resultado** de una evaluación heurística
 - Una **lista de problemas de usabilidad** que el evaluador ha encontrado en el diseño de la interfaz del sistema, junto a recomendaciones de resolución. **Cada problema se puede referir a uno o más heurísticos.**
 - Una **valoración acerca de la gravedad** de cada problema detectado, [URL](#), (posibles valores 0,1,2,3,4)
 - Gravedad: combinación de frecuencia, impacto y persistencia
- **Criterios** para determinar el cumplimiento de un heurístico:
 1. **Sencillo.** Responder en una escala de posibles valores sobre el cumplimiento del heurístico
 2. **Elaborado.** Plantear para cada heurístico una lista de comprobación formada por preguntas y posibles respuestas (respuesta libre, número, ...) y decidir los valores que determinan el cumplimiento del heurístico

7.2 Técnicas de Evaluación de la Usabilidad. Recorrido cognitivo

- Cualquier usuario prefiere aprender el uso de un software para una tarea particular **explorando las posibilidades de ejecución**, más que siguiendo un estricto orden de ejecución de pasos en cada tarea
- El recorrido cognitivo implica a un grupo de evaluadores que examinan la interfaz realizando un conjunto de tareas y evaluando su comprensión y facilidad de aprendizaje
 - Cada evaluador ejecuta los pasos de una tarea e infiere usando exploración **cómo un usuario final ejecutaría cada paso de la tarea**
- Esta técnica es más apropiada en la **etapa de diseño**

7.2 Técnicas de Evaluación de la Usabilidad. Inspección de estándares

- Ya tratado en Tema 3 (Calidad)
- **Clasificación**
 - **Estándares de iure**
 - Generados por **comités** de estatus legal (ISO, ANSI, IEEE, ...)
 - **Estándares de facto**
 - A partir de productos con éxito en mercado (p.ej. sistema *X-Windows*)
- Un evaluador realiza una inspección de la interfaz para comprobar el cumplimiento del estándar, sin fijarse en la funcionalidad de la aplicación
- Esta técnica es más apropiada en la **etapa de producto final**

7.2 Técnicas de Evaluación de la Usabilidad

- Métodos de evaluación de la usabilidad según criterio “técnica de comprobación”
 - Inspección
 - **INDAGACIÓN. Técnicas:**
 1. Observación de campo (*field observation*)
 2. Grupo de discusión dirigido (*focus group*)
 3. Entrevistas
 4. Cuestionarios
 5. Registro de uso
 - Test



Interacción con usuarios finales, pero sin tests

7.2 Técnicas de Evaluación de la Usabilidad. Observación de campo

- Ya tratado en sección “Análisis” (denominado “Análisis Etnográfico”)
 - Usable en **etapas de análisis y de producto final**
 - **Visitar, observar y recoger** datos para usuarios representativos en su lugar de trabajo, con un posterior análisis de los datos



7.2 Técnicas de Evaluación de la Usabilidad

- **Grupo de discusión dirigido** (*focus group*)
 - Ya tratado en sección “Análisis”
 - Grupo de 6-9 personas (usuarios y/o implicados) para tratar de **un aspecto del sistema**
- **Entrevistas y cuestionarios**
 - Ya tratado en sección “Análisis”. Válidos en cualquier fase del ciclo de vida.
 - Los cuestionarios son menos flexible que entrevistas, pero mejoran la cobertura de usuarios y el rigor en tratamiento de datos

7.2 Técnicas de Evaluación de la Usabilidad. Registro de uso

- Una **aplicación recoge logs del usuario** y el proceso es transparente para el mismo
- **Características**
 - Método económico y amplia cobertura
 - Usuario en entorno habitual
 - No necesaria presencia física usuarios
 - Tratamiento estadístico
 - Usable en etapa de producto final

7.2 Técnicas de Evaluación de la Usabilidad

- Métodos de evaluación de la usabilidad según criterio “técnica de comprobación”
 - Inspección
 - Indagación
 - **TEST. Técnicas:**
 1. Medida de las prestaciones
 2. Pensar en voz alta (*thinking aloud*)
 3. Ordenación de tarjetas (*card sorting*)



Basadas en experimentos/pruebas/tests con usuarios finales

7.2 Técnicas de Evaluación de la Usabilidad. Medida de las prestaciones

- Medidas (**métricas**) de:
 - **Rendimiento** (*performance*)
 - Medidas cuantitativas
 - Relación con {*effectiveness*, *efficiency*}
 - **Aspectos subjetivos** de usabilidad
 - Percepción (cuantitativas o cualitativas)
 - Relación con {*satisfaction*}
- El objetivo de la evaluación es **usabilidad y no funcionalidad**
- Criterios para la **elección de tareas**
 - Tareas con inherentes problemas de usabilidad
 - Tareas habituales
 - Tareas derivadas (p.ej. tareas de recuperación después de un error)

Usability goals (ISO 9241-11)



7.2 Técnicas de Evaluación de la Usabilidad. Medida de las prestaciones

- Ejemplos de medidas de **rendimiento**
 - Tiempo
 - Finalizar tarea
 - Navegación en menús
 - Consultas de ayuda
 - Recuperación de errores
 - Número
 - Errores
 - Opciones de menú
 - Teclas de función
 - Llamadas a la ayuda



7.2 Técnicas de Evaluación de la Usabilidad. Medida de las prestaciones

- Ejemplo de medida de **rendimiento cualitativo** a partir de datos cuantitativos

<i>Medidas de tiempo (tarea 1)</i>	<i>Excelente</i>	<i>Aceptable</i>	<i>Inaceptable</i>
Finalización	< 3 min	3-5 min	> 5 min
Recuperación de un error	0	< 1 min	> 1 min
Ayuda en línea	< 2 min	2-3 min	> 3 min

7.2 Técnicas de Evaluación de la Usabilidad. Medida de las prestaciones

- Ejemplos de medidas de **aspectos subjetivos**
 - Estimaciones de facilidad de:
 - Uso
 - Aprender
 - Ejecución de tarea
 - Instalación
 - Encontrar información
 - Preferencias
 - Versión previa
 - Producto de la competencia
 - Predicciones de comportamiento
 - ¿Comprará el producto?

7.2 Técnicas de Evaluación de la Usabilidad. Pensar en voz alta

- Nielsen, 1993
- Se proporciona a los usuarios un sistema a probar y un conjunto de tareas. A continuación, los usuarios se **expresan en voz alta** mientras interaccionan con el sistema.
- Permite a los evaluadores comprender **el modelo mental y la terminología del usuario**
- **Ventajas**
 - Simple y económico
 - Puede hacerse en cualquier fase del ciclo de vida
 - Proporciona ideas muy útiles de la usabilidad del sistema

7.2 Técnicas de Evaluación de la Usabilidad. Ordenación de tarjetas

- Ya tratado en sección “Análisis”
- Pertenece al diseño de la Arquitectura de la Información
- Permite agrupaciones de una larga lista de elementos no relacionados
- Fácil de entender y barata
- Clasificación
 - Ordenación abierta (sin grupos iniciales)
 - Ordenación cerrada (con grupos iniciales con nombre)
- Más habitual en entornos web
- Existen técnicas de evaluación comparativa para ordenación de tarjetas (detallado al final del tema)

7.2 Técnicas de Evaluación de la Usabilidad

- Métodos de evaluación en el ciclo de vida

Método	Análisis	Diseño	Prod. Final
Evaluación heurística		x	x
Recorrido cognitivo		x	x
Inspección de estándares			x
Observación de campo	x		x
Entrevistas, Cuestionarios	x	x	x
Medida de prestaciones		x	x
Pensar en voz alta	x	x	x

7.2 Técnicas de Evaluación de la Usabilidad

- El **coste de evaluación de usabilidad** importa en la elección de un método de evaluación
 - Personal, tiempo, ...
- **Clasificación** (según coste)
 - Bajo
 - Evaluación heurística
 - Medio
 - Recorrido cognitivo
 - Observación de campo
 - Entrevistas
 - Cuestionarios
 - Alto
 - Grupo de discusión dirigido
 - Medida de prestaciones
 - Pensar en voz alta

7.3 Experimentación con Usuarios

- **Libros de referencia** ([URL1](#) 2013, [URL2](#) 2017)
- ***User experience (UX)***
 - Concepto relativo a todas las características de interacción de un usuario final con un producto o sistema, **no sólo a nivel de su usabilidad**, por ejemplo:
 - Atractivo, sugerente, entretenido, estético, adictivo, ...
- **Métricas de usabilidad**
 - Tiempo, tasas de éxito, número errores, satisfacción, ...
 - Técnicas de rastreo ocular, ...
- **Técnicas importantes** para análisis de datos de métricas
 - Intervalos de confianza
 - Comparación entre grupos
 - Variables independientes (edad, género, prototipo, ...)
 - Variables dependientes (resultados de las métricas: tiempo, éxito, ...)
 - Inferencia Estadística
 - Permite deducir datos generales a partir de muestras

7.3 Experimentación con Usuarios.

Factores importantes

- **Factores importantes para el diseño de un estudio de usabilidad**
 - 1) *Selección de participantes*
 - 2) *Número de participantes*
 - 3) *Muestras de un grupo de participantes o de varios grupos*
 - 4) *Compensación*
 - 5) *Identificación de variables independientes y dependientes*
- **1) Selección de participantes**
 - **Factores:** disponibilidad, coste, objetivos del estudio, ...
 - **Criterios**
 - Representatividad
 - Agrupaciones
 - Experiencia en un dominio (principiante, intermedio, experto)
 - Factores demográficos (género, edad, lugar, ...)
 - Estrategia de selección
 - Muestra aleatoria
 - Muestra estratificada (por ejemplo 50% hombres y 50% mujeres)
 - Publicidad (buscar voluntarios, anunciar, ...)

7.3 Experimentación con Usuarios.

Factores importantes

- **2) Número de participantes**

- Depende del objetivo del estudio y la tolerancia de errores

- **Ejemplos**

- Identificar puntos más importantes de usabilidad como parte de un proceso iterativo de diseño
 - 3 o 4 participantes representativos
- Muchas tareas o varias partes del producto
 - Un número mayor

- **Tolerancia de errores**

- Si 8 de 10 usuarios completan una tarea, se puede asumir una media de 80% con un intervalo de confianza de mayor longitud que si lo hacen 40 de 50 usuarios, *Tab. a*
- Concepto de intervalo de confianza de 95%
 - Para el 95% de veces que se calcule un intervalo de confianza, la media real estará dentro del intervalo de confianza

7.3 Experimentación con Usuarios.

Factores importantes

Éxitos	Participantes	Int. Conf. 95%, a	Int. Conf. 95%, b
4	5	36%	98%
8	10	48%	95%
16	20	58%	95%
24	30	62%	91%
40	50	67%	89%
80	100	71%	86%

Tab. a

Se está usando el método de ajuste de Wald para el cálculo del intervalo de confianza del promedio de una distribución Bernoulli

7.3 Experimentación con Usuarios.

Factores importantes

- **3) Muestras de un grupo de participantes o de varios grupos**
 - Existen principalmente **dos diseños**
 - **Medidas intra-sujetos (*within-subjects*) o relacionadas.** Cada participante realiza varias pruebas, p.ej. tiempos para diferentes diseños de un prototipo
 - **Medidas inter-sujetos (*between-subjects*) o independientes.** Cada participante realiza una sola prueba y los participantes pertenecen a grupos independientes, p.ej. tiempo de tarea para hombres o mujeres.
 - **Comparación de los diseños**
 - **A) Medidas intra-sujetos**
 - **Ventajas**
 - Se necesitan menos participantes
 - Habitual para medir facilidad de aprendizaje (p.ej. usando como métricas tiempos o número de errores). En otros casos aparecen algunos **inconvenientes**:
 - Al compararse un participante consigo mismo, pueden existir factores de acumulación (*carryover*) no deseables: fatiga, aprendizaje adquirido en prueba anterior, etc. En este caso, se podrían buscar alternativas de compensación (*counterbalancing*) para compensar los factores de acumulación.

7.3 Experimentación con Usuarios.

Factores importantes

- **B) Medidas inter-sujetos**
 - El número de participantes puede ser diferente en cada grupo.
 - Ejemplo: diferencias de satisfacción o de tiempos entre principiantes y expertos o entre diferentes diseños de un prototipo
 - **Inconvenientes**
 - Se necesitan más participantes
 - **Ventajas**
 - Eliminación de efectos de acumulación
- **C) Diseños mixtos**
 - Repetir medidas para cada usuario de cada grupo, p.ej. tiempos de compleción de diferentes tareas para hombres y mujeres

7.3 Experimentación con Usuarios.

Factores importantes

- **4) Compensación** (en medidas intra-sujetos)
 - Permite evitar efectos de acumulación derivados del orden de tareas
 - Ejemplo:
 - Cambiar el orden para evitar repeticiones horizontales/verticales. Por ejemplo para 4 participantes y 4 tareas sin repeticiones en cada fila o columna:
 - ((T1,T2,T3,T4),(T3,T1,T4,T2),(T2,T4,T1,T3),(T4,T3,T2,T1))
 - El cambio de orden en tareas no tendría sentido para:
 - Tareas no relacionadas (y por tanto sin poder aprender de una a otra)
 - Tareas con un orden natural de ejecución
- **5) Identificación de variables independientes y dependientes**
 - Ejemplo: rendimiento para “hombres-mujeres” o para “principiantes-expertos” o para “diseño1-diseño2”
 - Variables independientes
 - Género, experiencia, diseño
 - Variables dependientes
 - Rendimiento (tasa de compleción, tiempo, etc.)
 - Las variables dependientes son los resultados y, por tanto, las métricas

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- Referencia ([URL](#), capítulo 2)
- Tipos de datos
 - **Nominal** (*nominal*)
 - No ordenados, *ej: {Windows, Mac}*
 - **Ordinal** (*ordinal*)
 - Ordenados, *ej: {mal, regular, bien}*
 - **Continuo relativo** (*interval*)
 - Sólo importan las diferencias
(*ej: temperaturas, ej: satisfacción-en-escala-1-5*)
 - **Continuo absoluto** (*ratio*)
 - Existe un cero absoluto (*ej: tiempos, edad*)
- Elementos de **Estadística Descriptiva**
 - **Inferencia Estadística**
 - Generalizar resultados a partir de muestras
 - *Excel (>=Excel 2010)*
 - Funciones (`INTERVALO.CONFIANZA.T`, `PRUEBA.T`, ...), [URL](#)

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- **Intervalos de confianza para promedios**
 - Ejemplo (*confianza 95%*)
 - *Con un 95% de confianza la media real está en el intervalo $[mediaMuestral - L, mediaMuestral + L]$*
 - Fichero `dsi-estadistica.xls`. Pestañas `interv-conf` e `interv-conf-bernoulli`
 - Función `INTERVALO.CONFIANZA.T`
 - Los resultados de promedios de métricas normalmente se acompañan de sus intervalos de confianza
 - **Cálculos dependiendo de la distribución inicial**
 - Distribución Normal → considerando una t-student auxiliar
 - Distribución Bernoulli ($p=probabilidad\ éxito$) → utilizando el método de ajuste de Wald
 - Para una distribución arbitraria → considerando que el promedio es normal para N grande (teorema del límite central)
 - N=número-medidas. Grande={ $N>30$ }, por simplicidad.

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- **Comparaciones de medias**
 - Valores de métricas en variables dependientes para diferentes valores de variables independientes
 - Se usan técnicas estadísticas de **contraste de hipótesis**, con la hipótesis nula “las-medias-son-iguales”
 - Si los intervalos de confianza son disjuntos, no es necesario el contraste de hipótesis, y se dice que **es estadísticamente significativo** que las medias son diferentes
- **Contraste de hipótesis (*significance tests*)**
 - *¿Cuál es el tamaño de la muestra?*. Normalmente, si es menor que 30 se usan “*Pruebas t o chi-cuadrado*”, y si es mayor se realizan “*Pruebas z*”
 - *¿Cuántas muestras se comparan?*. Si se comparan 2 muestras, se realiza una “*Prueba t o chi-cuadrado*”, y si hay más de 2 se realizan ANOVA (*analysis of variance*) o pruebas chi-cuadrado
- **Variantes**
 - 1) Medidas inter-sujetos (prueba t)
 - 2) Medidas intra-sujetos (prueba t)
 - 3) Medidas inter-sujetos (chi-cuadrado)

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- **1) Medidas inter-sujetos (prueba t)**
 - Ejemplo: *tiempo de una tarea para hombres y mujeres*
 - *Variable-independiente=género*. Realizada para 8 mujeres y 10 hombres.
 - Cada muestra tiene un valor diferente de una variable independiente
 - Cada muestra puede contener un **número diferente de participantes**
 - Se suponen 2 distribuciones normales con la misma varianza
 - *Hipótesis nula*: las medias son iguales
 - “*Prueba t de Student*”, para rechazar la hipótesis nula
 - Confianza: 95%, Significación: 5%
 - Si se rechaza la hipótesis nula, se dice que **es estadísticamente significativo** que las medias son diferentes
 - Fichero `dsi-estadistica.xls`. Pestaña `prueba-t-inter-suj`
 - Función `PRUEBA.T`

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- **2) Medidas intra-sujetos (prueba t)**
 - **Cada participante repite un experimento** para 2 valores diferentes de una variable independiente
 - Ejemplo: *tiempo de una tarea para dos tipos de interfaz (táctil y no táctil)*
 - *Variable-independiente=tipo-interfaz*. Realizada para 5 participantes.
 - Se suponen distribución normal para la diferencia de valores
 - *Hipótesis nula*: las media de la diferencia es 0
 - “*Prueba t de Student*”, para rechazar la hipótesis nula
 - Confianza: 95%, Significación: 5%
 - Fichero `dsi-estadistica.xls`.
 - Pestaña `prueba-t-intra-suj`
 - Función `PRUEBA.T`

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- **3) Medidas inter-sujetos (prueba chi-cuadrado)**
 - Usado para datos nominales y ordinales en las variables dependientes
 - Ejemplo: *compleción con éxito de tareas*
 - “Prueba chi-cuadrado”
 - Es la más común de las pruebas no paramétricas (distribución inicial no conocida)
 - Ejemplo (3 grupos)
 - *Compleción de tareas para {principiante,intermedio,experto}*
 - Fichero `dsi-estadistica.xls`. Pestaña `chi-cuadrado`
 - Función `PRUEBA.CHI`
- **Relación entre variables**
 - **Coeficiente de correlación**
 - Fichero `dsi-estadistica.xls`. Pestaña `coef-correlac`
 - Función `COEF.DE.CORREL`

7.3 Experimentación con Usuarios.

Estadística Descriptiva

- **Presentación gráfica de datos**
 - **Gráficos de barras**
 - Para *valores-continuos/valores-discretos*, ej: *tasa-éxito/tarea*, *Fig. a*
 - **Gráficos de líneas**
 - Para *valores-continuos/valores-continuos*, ej: *tasa-éxito/edad*
 - **Gráficos de puntos**
 - Para *valores-continuos/valores-continuos*, ej.: *página-atractiva/tamaño-de-mayor-imagen-de-página*, *Fig. b*
 - Útil mostrar la línea de tendencia y el coeficiente de correlación
 - **Gráficos de tarta**
 - Proporciones, ej: porcentaje de páginas con errores de accesibilidad para 0 errores, 1-2 errores, etc., *Fig. c*
 - **Gráficos de barras apiladas**
 - Son varios gráficos tarta agrupados en un gráfico de barras, ej: porcentaje de participantes que aciertan/fracasan para {encontrar información de contacto, introducir comentarios, ...}, *Fig. d*
- Los **intervalos de confianza** aparecen como **barras de error** para valores promedios (p.ej. en *Fig. a*)

7.3 Experimentación con Usuarios.

Estadística Descriptiva

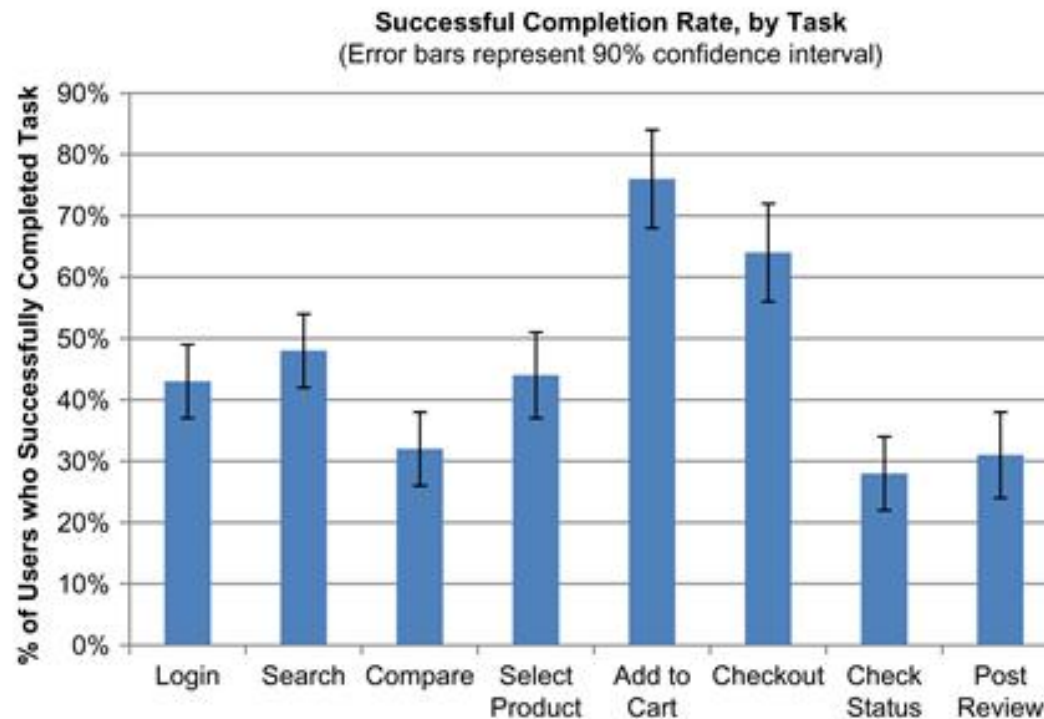


Fig. a

7.3 Experimentación con Usuarios.

Estadística Descriptiva

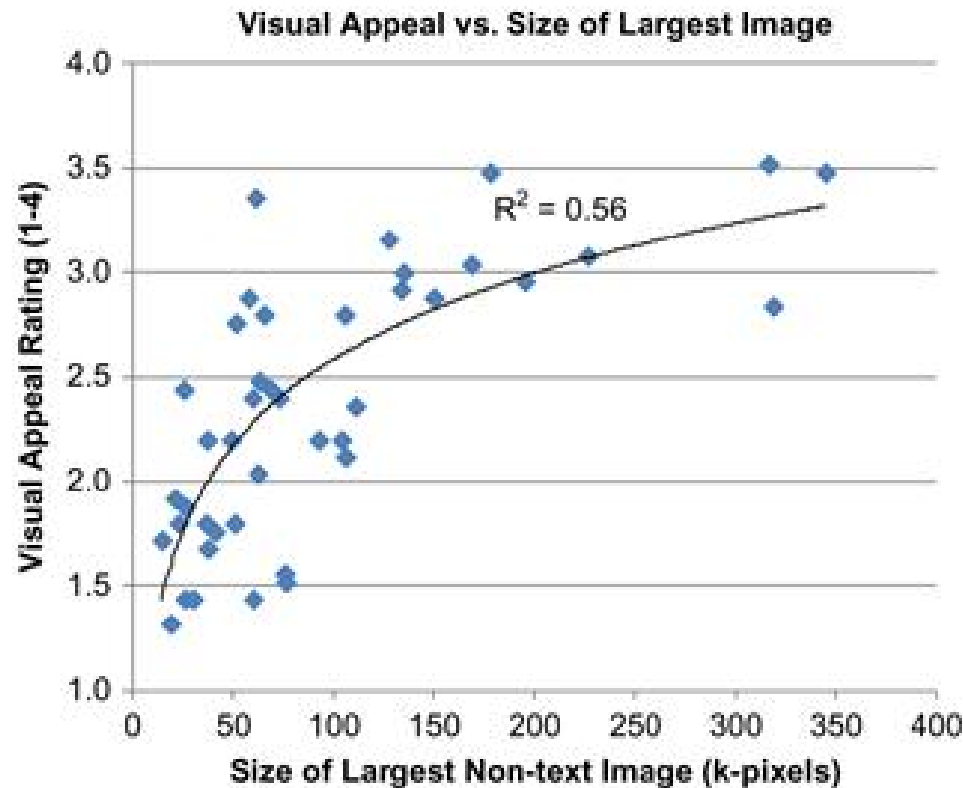


Fig. b

7.3 Experimentación con Usuarios.

Estadística Descriptiva

% of Pages with Accessibility Errors

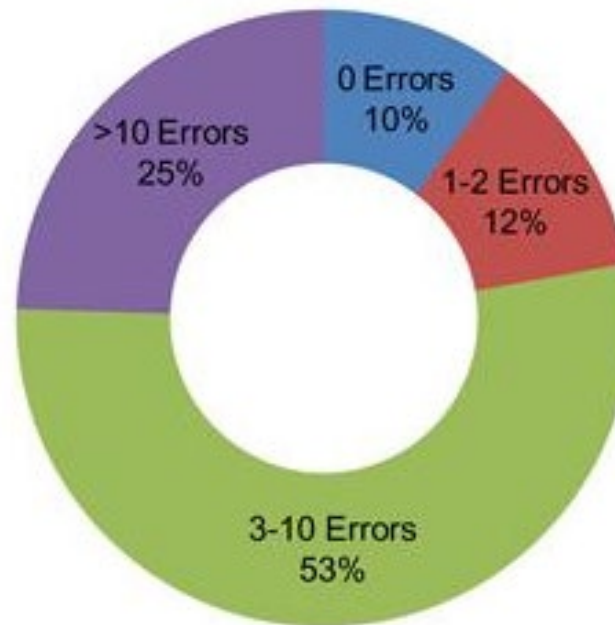


Fig. c

7.3 Experimentación con Usuarios.

Estadística Descriptiva

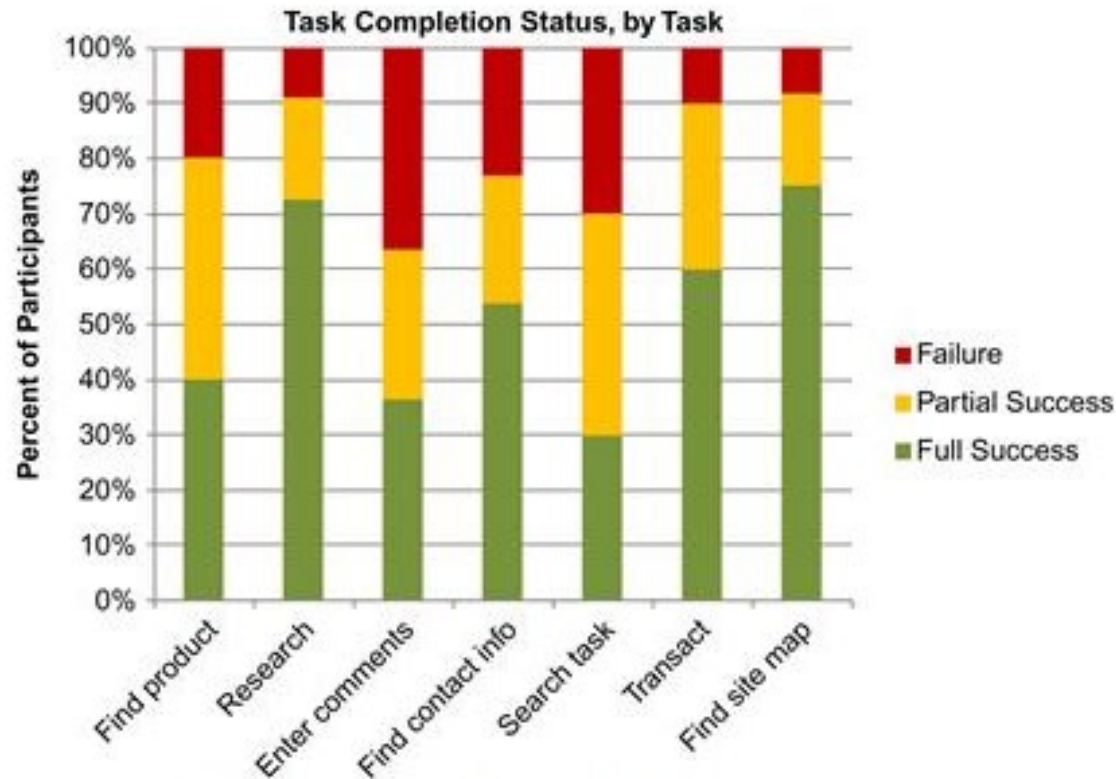


Fig. d

7.3 Experimentación con Usuarios.

Planificación de un estudio de usabilidad

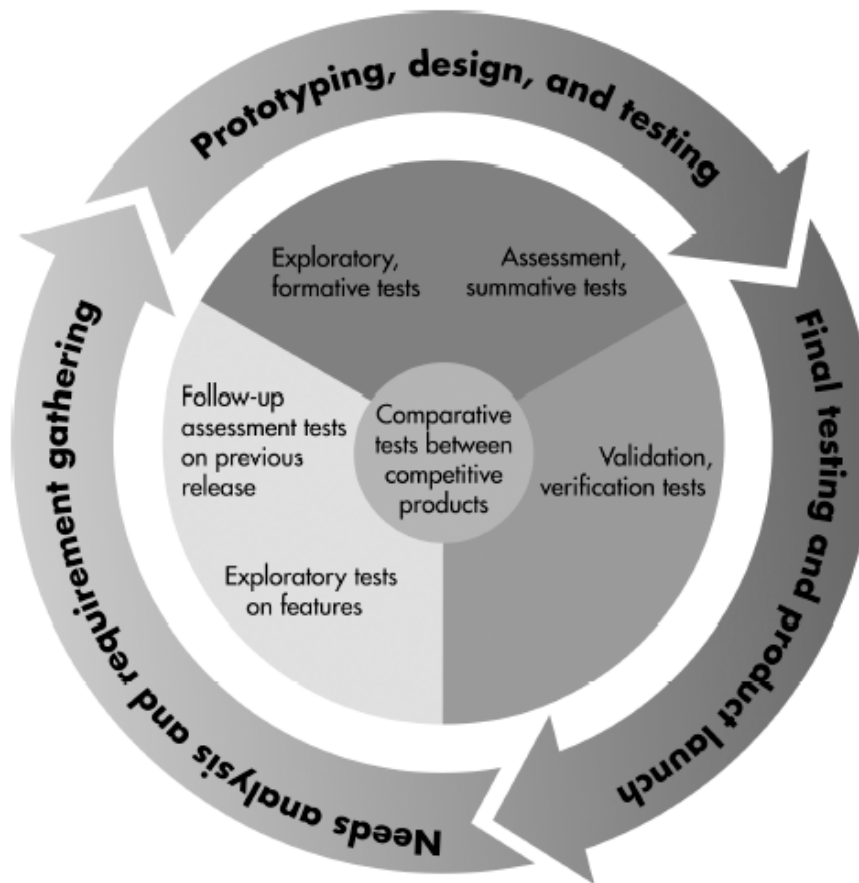
- La planificación de una evaluación/estudio de usabilidad debe determinar:
 - **1) Objetivos**
 - ¿Fase del ciclo de vida y tipo de pruebas?. ¿Qué se va a evaluar (identificar el escenario)?. ¿Cuáles son las métricas más apropiadas?
 - **2) Tecnología de recogida de datos, presupuesto, tiempo, etc.**
- **1) Objetivos del estudio**
 - Identificar la fase del ciclo de vida y el tipo de pruebas (*Fig a; [URL](#)*, capítulo 3). Por ejemplo en fase de diseño:
 - **Pruebas formativas (*formative*) de usabilidad**
 - Sobre versiones iterativas del mismo producto. Más común en primeras fases del ciclo de vida.
 - **Pruebas sumativas (*summative*) de usabilidad**
 - Sobre prototipos más completos

7.3 Experimentación con Usuarios.

Planificación de un estudio de usabilidad

Fig. a

Usability testing throughout the product lifecycle



7.3 Experimentación con Usuarios.

Planificación de un estudio de usabilidad

Ten common usability study scenarios and the metrics that may be most appropriate for each

Usability Study Scenario	Task Success	Task Time	Errors	Efficiency	Learn-ability	Issues-based Metrics	Self-reported Metrics	Behavioral & Physiological Metrics	Combined & Comparative Metrics	Live Website Metrics	Card-Sorting Data
1. Completing a transaction	X			X		X	X			X	
2. Comparing products	X			X			X		X		
3. Evaluating frequent use of the same product	X	X		X	X		X				
4. Evaluating navigation and/or information architecture	X		X	X							X
5. Increasing awareness							X	X		X	
6. Problem discovery						X	X				
7. Maximizing usability for a critical product	X		X	X							
8. Creating an overall positive user experience							X	X			
9. Evaluating the impact of subtle changes										X	
10. Comparing alternative designs	X	X				X	X		X		

Métricas apropiadas para diferentes escenarios ([URL](#), capítulo 3.3)

7.3 Experimentación con Usuarios.

Planificación de un estudio de usabilidad

- Identificar las métricas más apropiadas para el escenario:
 - Rendimiento
 - Tasa de éxito, tiempo, esfuerzo (número de clicks, esfuerzo cognitivo, etc.), errores, tiempo para dominarlo (*learnability*), facilidad de aprendizaje de tareas básicas (*ease of learning*), ...
 - Satisfacción
 - Fácil de usar, confuso, supera expectativas, ...
- **2) Tecnología de recogida de datos, presupuesto, tiempo, etc.**
 - Tiempos
 - Mitad en la organización de la ejecución, identificando y validando tareas, creando preguntas y escalas, seleccionando participantes, ...
 - Mitad en recogida y análisis de datos
 - Participantes
 - Laboratorio/On-line
 - Laboratorio (4-10 participantes)
 - Pruebas formativas (6-8 participantes) y sumativas (10-50)

7.3 Experimentación con Usuarios.

Planificación de un estudio de usabilidad

- **Resumen (en el contexto de la práctica 3)**
 - En fase de análisis, se determinan los **atributos de usabilidad (métricas con valores objetivo)** a lo largo de su ciclo de vida
 - En cualquier punto del ciclo de vida, por ejemplo en fase de diseño, se puede realizar una **evaluación de usabilidad usando pruebas con usuarios finales**, y basada en una **planificación**
 - Contenido de la planificación
 - Objetivos del estudio y tecnologías
 - Factores importantes del “diseño de un estudio de usabilidad” (selección de participantes, número de participantes, ...)
 - Pruebas → medidas de las prestaciones, pensar en voz alta, ...
 - Ejemplo de planificación → Un sistema web de reservas de hotel (ejemplo en fichero `Hdotcom_test_plan.pdf`; capítulo 5 de [URL](#), extraído de [URL](#))

7.3 Experimentación con Usuarios.

Métricas de rendimiento

- **Usabilidad** (eficacia + eficiencia + satisfacción)
 - **Rendimiento** (eficacia + eficiencia)
 - *Effective vs efficient → Being effective is about doing the **right things**, while being efficient is about doing the things in the **right manner***
- Aproximadamente 8 observaciones pueden obtener intervalos de confianza razonables
- Principales métricas de rendimiento
 - 1) **Éxito de tareas**
 - 2) **Tiempo de tareas**
 - 3) **Errores de tareas**
 - 4) **Eficiencia de tareas**
 - 5) **Facilidad de dominio**

7.3 Experimentación con Usuarios.

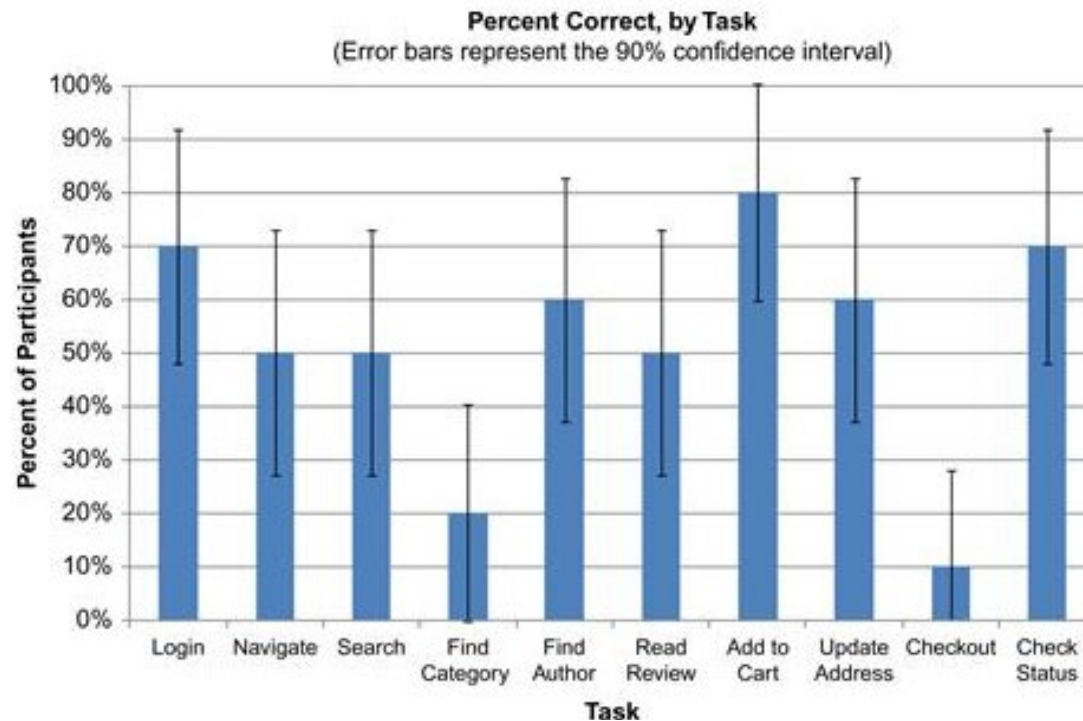
Métricas de rendimiento

- **1) Éxito de tareas**

- Es la más usada (compleción de una o más tareas)
- **Éxito binario**
 - **Gráfico de barras** (*tasa-éxito/tarea*) con intervalos de confianza como barras de error, p.ej. *Fig. a*
 - Posibles **comparaciones de medias** usando contraste de hipótesis con *pruebas chi-cuadrado* en el caso sumativo, como por ejemplo:
 - 2 tareas
 - {usuario-habitual, usuario-no-habitual}
 - {experiencia-previa-con-producto, sin-experiencia-previa}
 - {grupo1-edad, grupo2-edad}
- **Éxito con niveles**
 - Son **datos ordinales** y se presentan como medias en un **gráfico de barras apiladas** (*nivel-éxito/tarea*), p.ej. *Fig b*

7.3 Experimentación con Usuarios.

Métricas de rendimiento

*Fig. a**Task success rates, including a 90% confidence interval for each task*

7.3 Experimentación con Usuarios.

Métricas de rendimiento

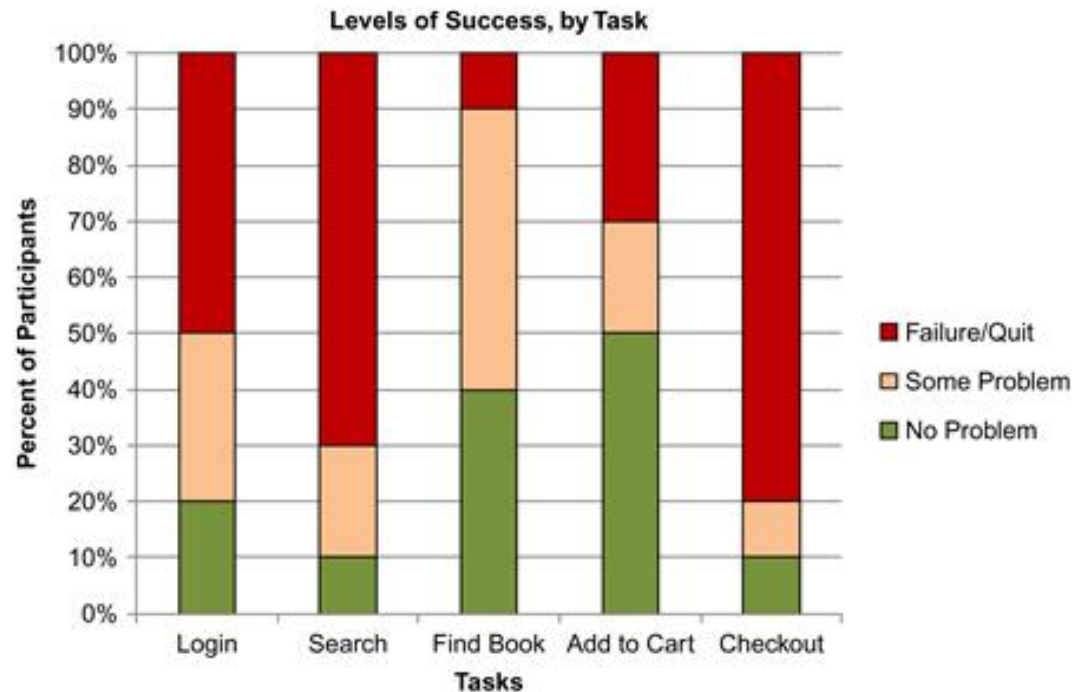


Fig. b

Stacked bar chart showing different levels of success based on task completion.

7.3 Experimentación con Usuarios.

Métricas de rendimiento

• 2) Tiempo de tareas

- Normalmente es **medida de eficiencia** (p.ej. tiempo para hacer una reserva en la web)
 - En algunos casos, **puede no ser un factor de usabilidad**
 - Juegos en los que el tiempo consumido no es el objetivo
 - Aprendizaje
- **Gráfico de barras** (*tiempo/tarea*) con intervalos de confianza como barras de error, p.ej. *Fig. a*
- Posibles **comparaciones de medias** con *pruebas t* en el caso sumativo (2 tareas, etc.)
- Prácticas habituales
 - Excluir **casos atípicos** (*outliers*)
 - Considerar solo **tareas con éxito**
 - Pero se considerarían todas si se tratara de obtener una visión global de experiencia de usuario

7.3 Experimentación con Usuarios.

Métricas de rendimiento

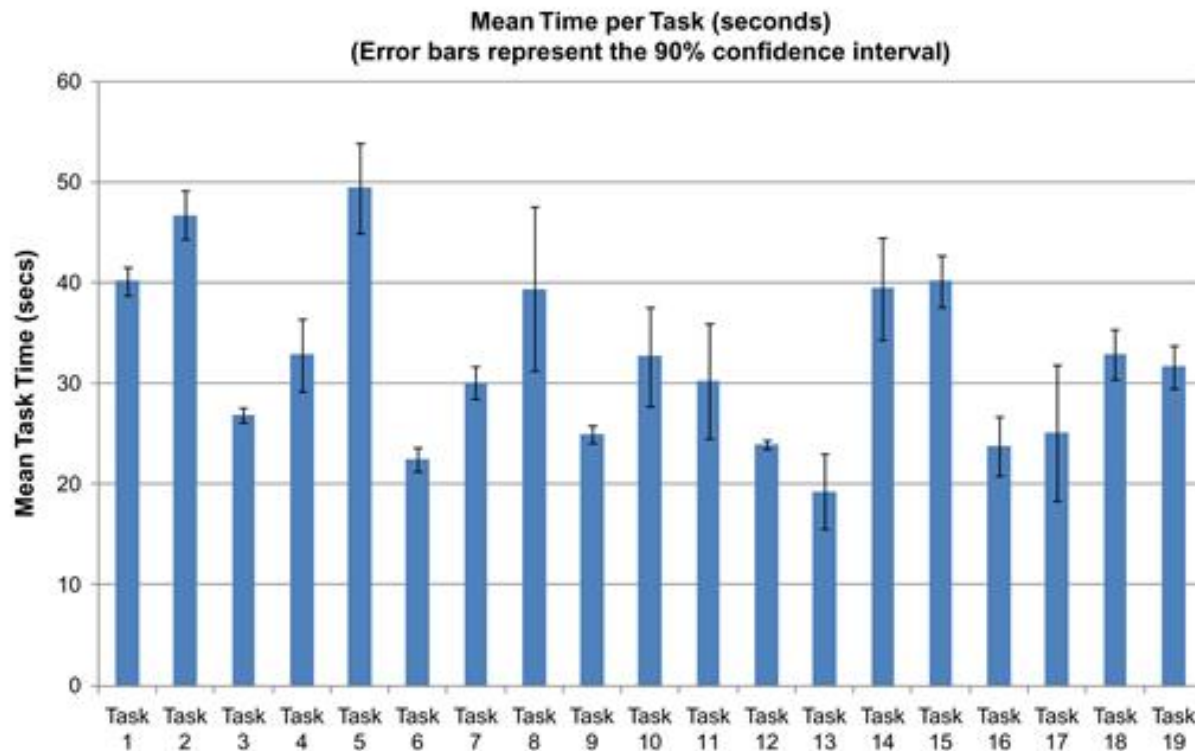


Fig. a

Mean time on task, in seconds, for 19 tasks. Error bars represent a 90% confidence interval. These data are from an online study of a prototype website.

7.3 Experimentación con Usuarios.

Métricas de rendimiento

- Posible uso de **rangos** de posibles valores del tiempo de tarea
 - Gráfico *porcentaje-usuarios-en-rango/tarea*, p.ej. *Fig. b*

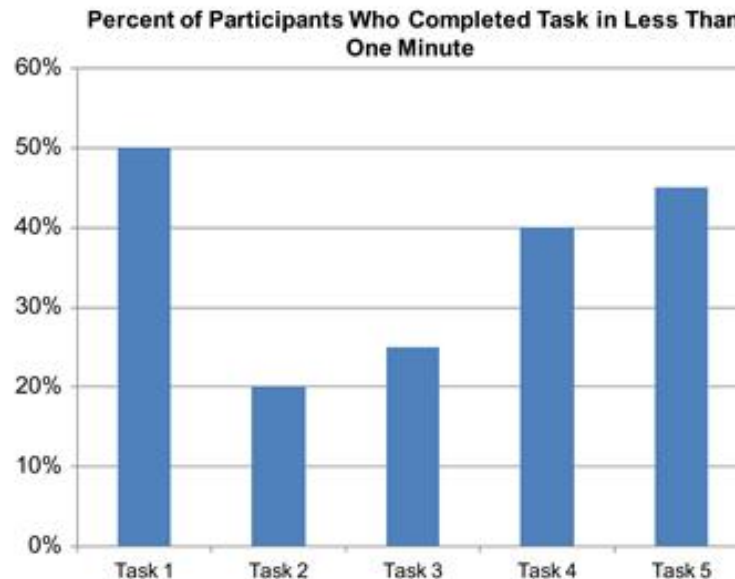


Fig. b

An example showing the percentage of users who completed each task in less than 1 minute.

7.3 Experimentación con Usuarios.

Métricas de rendimiento

- **3) Errores de tareas**
 - **Acciones incorrectas** del usuario, p.ej. contraseña errónea
 - En muchas otras ocasiones, los errores son consecuencia de **problemas de usabilidad**
 - Las **medidas de errores son útiles** en casos como:
 - Si supone un **coste significativo** (p.ej. mejora de servicios de atención al cliente)
 - Si conduce a **fracaso en tareas** (p.ej. comprar un producto por error en un sitio web)

7.3 Experimentación con Usuarios.

Métricas de rendimiento

- **4) Eficiencia de tareas**

- Eficiencia → *¿cuánto cuesta realizarlas? ¿se han realizado de la mejor forma posible?*
 - Se suelen considerar sólo **tareas con éxito**
- Diferentes **medidas de eficiencia** en una tarea
 - 1) Tiempo (tratado en apartado 2)
 - 2) Esfuerzo
 - 3) Desorientación
 - 4) Combinación de éxito y tiempo
- **2) Esfuerzo** para completar una tarea
 - **Cognitivo** (p.ej: “encontrar un link en una página web”, “decidir si hacer click” e “interpretar el resultado de la acción”)
 - **Físico** (p.ej: “mover el ratón”, “introducir texto”, etc.)
 - Ejemplo: medida del número de *clicks*

7.3 Experimentación con Usuarios.

Métricas de rendimiento

- 3) **Desorientación** (*lostness*). En entornos web, es común el uso de la métrica L (p.ej. *Fig. a* $\rightarrow R=3, N=6, S=9$)
 - $L = \sqrt{[(N/S-1)^2 + (R/N-1)^2]} \quad (R \leq N \leq S)$
 - N=número de páginas web diferentes visitadas en una tarea
 - S=número de páginas web visitadas, incluyendo las revisitadas
 - R=número mínimo óptimo de páginas visitadas en una tarea
 - Camino perfecto ($R=N=S \rightarrow L=0$). Algunos estudios han estimado $L>0.5$ como región crítica de desorientación.
- 4) Otra medida de eficiencia, en el caso de que cada usuario realice varias tareas, es una **combinación de éxito-de-tareas y tiempo-de-tareas** (p.ej. *Fig. b*), calculando la división:
 - *Métrica = número-tareas-con-éxito / tiempo-de-tareas-con-éxito*
 - A mayores valores de la métrica, mayor eficiencia

7.3 Experimentación con Usuarios.

Métricas de rendimiento

Visited pages: {Home,B,Home,C,C2,C,C3,C,C1}

$$L = \text{sqrt}[(6/9 - 1)^2 + (3/6 - 1)^2] = 0.60$$

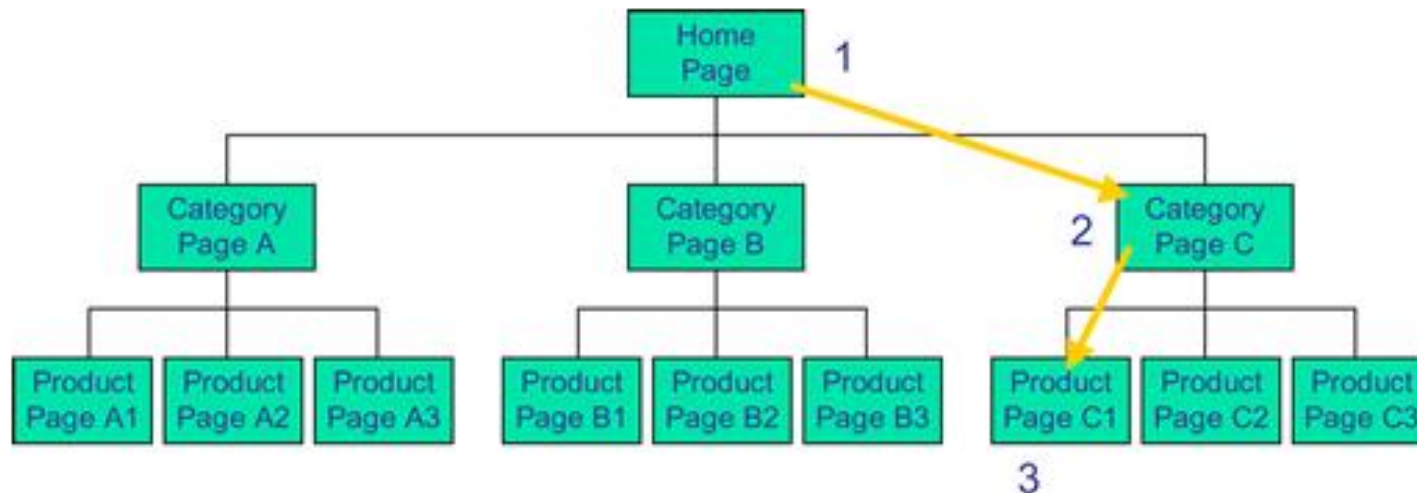


Fig. a

Optimum number of steps (three) to accomplish a task that involves finding a target item on Product Page C1 starting from the home page.

7.3 Experimentación con Usuarios.

Métricas de rendimiento

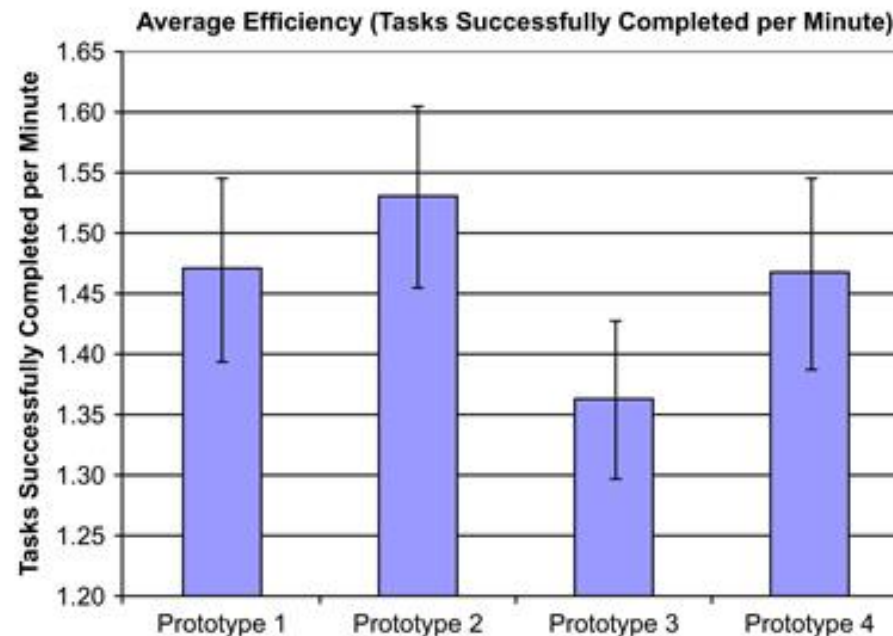


Fig. b

Average number of tasks completed successfully per minute in an online study of four different prototypes of navigation for a website. Over 200 participants attempted 20 tasks with each prototype. Participants using Prototype 2 were significantly more efficient (i.e., completed more tasks per minute) than those using Prototype 3.

7.3 Experimentación con Usuarios.

Métricas de rendimiento

- **5) Facilidad de dominio (*learnability*)**
 - Para medirla, los datos se recogen en **varias fases**, usando métricas de eficiencia para cada fase. Ejemplo de resultados:
 - **Gráfico de línea** *eficiencia/fase*, p.ej. *Fig. a*
 - Un ejemplo sería medir el tiempo necesario para dominar un proceso dependiendo de su frecuencia de ejecución (p.ej. diario/anual)

7.3 Experimentación con Usuarios.

Métricas de rendimiento

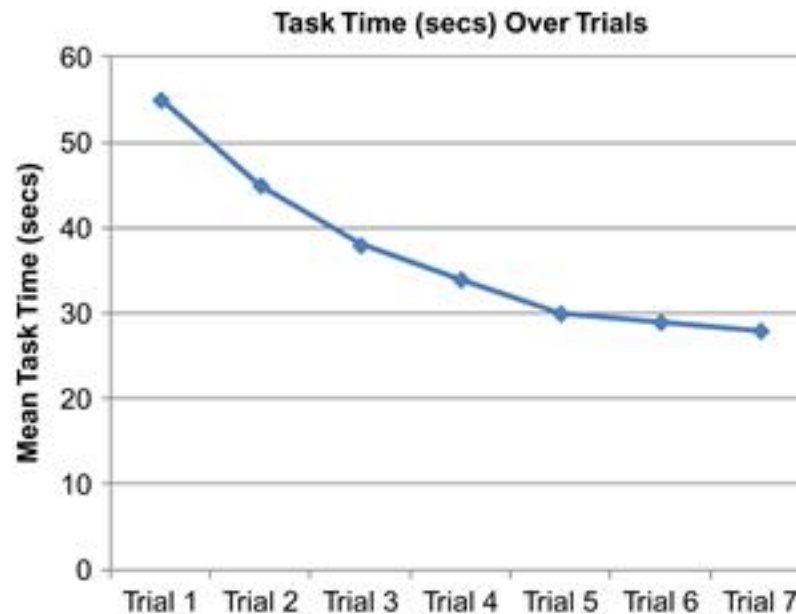


Fig. a

An example of how to present learnability data based on time on task.

7.3 Experimentación con Usuarios.

Métricas subjetivas

- **Métricas subjetivas**

- Las métricas de rendimiento se suelen considerar **objetivas**, a diferencia de las denominadas métricas **subjetivas** (*self-reported metrics*)
- Tienen por objetivo medir la **satisfacción** de los usuarios finales
- Los resultados de estas métricas, en ocasiones, **pueden ser más significativos** que los de métricas objetivas

- **Preguntas en recogida de datos**

- 1) Preguntas con **respuestas según una escala**
 - Escala de Likert
 - Escala diferencial semántica
- 2) Preguntas con **respuestas abiertas**

7.3 Experimentación con Usuarios.

Métricas subjetivas

- **Escala de Likert**

- Cada respuesta se basa en **5 opciones**:
 - *Totalmente en desacuerdo, En desacuerdo, Ni de acuerdo ni en desacuerdo, De acuerdo, Totalmente de acuerdo*
 - **Alternativamente**, de 7 opciones. Más común un **número impar**, para que exista un valor de neutralidad posible.
- La afirmación puede ser **positiva** (p.ej. “La terminología usada en la interfaz es clara”) o **negativa** (p.ej. “Encontré confusas las opciones de navegación”)

- **Escala diferencial semántica**

- Una respuesta se basa en una valoración, habitualmente con 5 posibilidades, para **un adjetivo y su contrario**. Ejemplo:
 - Caliente _ _ _ _ _ Frío

7.3 Experimentación con Usuarios.

Métricas subjetivas

- **Momento de recogida de datos**
 - Al final de una tarea (**post-tarea**)
 - Al final de un estudio o sesión (**post-sesión**). Más habitual.
- **Respuestas de usuarios**
 - 1) Orales
 - 2) Papel
 - 3) On-line
 - **Comparación**
 - 1) es el más fácil para el usuario
 - 2) y 3) convenientes en recogidas rápidas de datos
 - 3) normalmente son sitios de pago (*SnapSurveys.com*, etc.)

7.3 Experimentación con Usuarios.

Métricas subjetivas

- **Pautas generales** para escalas
 - Preguntar lo mismo de diferentes formas promediando al final los resultados, etc.
- **Análisis de datos**
 - Habitualmente, **asignando números** {1,2,3,4,5}
 - **Agrupaciones y frecuencias** para los grupos {1,2} y {4,5}
 - Más difícil para respuestas abiertas

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-tarea

- Técnicas
 - 1) Facilidad de uso
 - 2) Cuestionario ASQ (*After-Scenario Questionnaire*)
 - 3) Medida de expectativas
- 1) Facilidad de uso
 - Es la técnica **más sencilla y efectiva**. Se valora la *facilidad/dificultad* de **una tarea**. **Una única pregunta** con respuestas según:
 - Escala de Likert (*Esta tarea fue fácil de realizar*)
 - Escala diferencial semántica (*fácil/difícil*)
- 2) Cuestionario ASQ (ej. [URL](#))
 - ASQ, Jimm Lewis 1991, es aplicable a un escenario (**conjunto de tareas relacionadas entre sí**) y se basa en **3 preguntas** con respuestas según una escala de **7 opciones**:
 1. “Estoy satisfecho con la facilidad de terminar las tareas en este escenario” (*effectiveness, satisfaction*)
 2. “Estoy satisfecho con la cantidad de tiempo consumido en las tareas de este escenario” (*efficiency, satisfaction*)
 3. “Estoy satisfecho con la información de ayuda al realizar las tareas” (*satisfaction*)

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-tarea

- **3) Medida de expectativas** (Albet and Dixon, 2003)
 - Se valoran las reacciones subjetivas tras cada tarea, considerando para cada tarea **su dificultad en comparación con la dificultad esperada**
 - El usuario valora la dificultad esperada de las tareas antes de realizarlas, y también después de realizarlas, usando para las respuestas una escala de 7 puntos (*1=muy-difícil, 7=muy-fácil*)
 - Para cada tarea, se realiza un promedio para **dificultad esperada y dificultad real. Cuatro cuadrantes** definen las sugerencias para cada tarea. Ejemplo: *Fig. a*

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-tarea

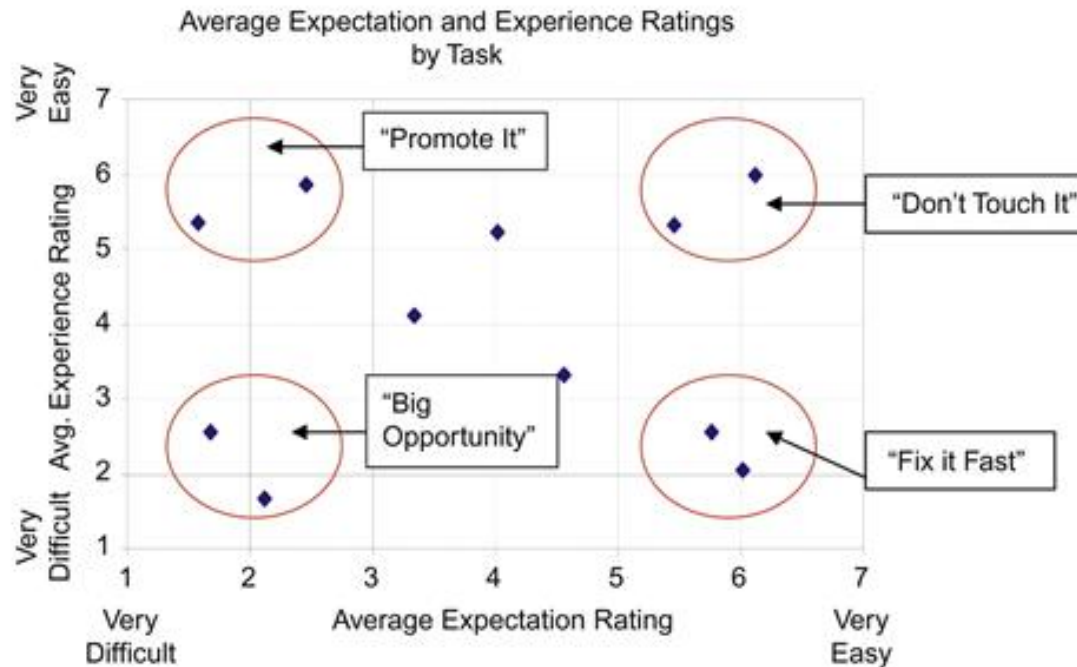


Fig. a

Comparison of average expectation ratings and average experience ratings for a set of tasks in a usability test. Which quadrants the tasks fall into can help you prioritize which tasks to focus on improving. Adapted from Albert and Dixon (2003); used with permission.

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

- Proporciona una **medida global de usabilidad subjetiva**
- Técnicas
 - 1) **Agregación de tareas individuales**
 - 2) **SUS (*System Usability Scale*)**
 - 3) **CSUQ (*Computer System Usability Questionnaire*)**
 - 4) **QUIS (*Questionnaire for User Interface Satisfaction*)**
 - 5) **USE (*Usefulness, Satisfaction, and Ease of Use Questionnaire*)**
- **Herramienta de generación de cuestionarios ([URL-ACM](#))**
 - Incluye QUIS, CSUQ, USE, ASQ
 - Se eligen los parámetros del tipo de cuestionario y se genera automáticamente un cuestionario web. Cada usuario rellena el formulario y se envía a una dirección de e-mail su respuesta.
- Las **conclusiones** que se obtienen de los resultados de cuestionarios se basan en estudios realizados sobre un amplio rango de casos

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

- **1) Agregación de tareas individuales**
 - Calcular el promedio de resultados para todas las tareas
- **2) SUS, [URL](#)**
 - Brooke, 1996, *Digital Equipment Corporation*.
 - **Método público**, pero tras su uso es obligatorio hacer mención a su origen
 - El **resultado global** se calcula con una fórmula y se expresa como un porcentaje
 - Un buen resultado SUS se ha estimado empíricamente en 70-80%
 - Son **10 afirmaciones** (5 positivas y 5 negativas), con respuestas según una escala diferencial semántica de **5 opciones** sobre el par *{totalmente-en-desacuerdo/totalmente-de-acuerdo}*, p.ej. Fig. a

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

	Strongly disagree					Strongly agree	
1. I think that I would like to use this system frequently	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		4
2. I found the system unnecessarily complex	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		1
3. I thought the system was easy to use	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		1
4. I think that I would need the support of a technical person to be able to use this system	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		4
5. I found the various functions in this system were well integrated	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		1
6. I thought there was too much inconsistency in this system	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		2
7. I would imagine that most people would learn to use this system very quickly	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		1
8. I found the system very cumbersome to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		1
9. I felt very confident using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>		4
10. I needed to learn a lot of things before I could get going with this system	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		3
Total = 22							
SUS Score = 22 * 2.5 = 55							

Fig. a

The System Usability Scale, developed by John Brooke at Digital Equipment Corporation and an example of scoring it.

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

- **3) CSUQ**

- Lewis (también autor de ASQ), 1995
- Son **19 afirmaciones positivas** con respuestas según una escala diferencial semántica de **5 opciones**, permitiendo la opción “sin-respuesta”, con valores extremos *{totalmente-en-desacuerdo, totalmente-de-acuerdo}*
- El **resultado global** que se obtiene está formado por **4 números**:
 - *{utilidad, calidad-información, calidad-interfaz, satisfacción-global}*

- **4) QUIS**

- Laboratorio de HCI, 1988, Univ. Maryland
- Son **27 afirmaciones** con respuestas según **escala 0-9**, permitiendo la opción “sin-respuesta”, divididas en 5 categorías:
 - *{reacción-global, pantalla, terminología, aprendizaje, capacidades-del-sistema}*

- **5) USE**

- Lund, 2001, *Fig. a*
- Son **30 afirmaciones positivas** divididas en **4 categorías**
 - *{utilidad, satisfacción, facilidad-de-uso, facilidad-de-aprendizaje}*
- El **resultado global** que se obtiene está formado por 4 números (uno por cada categoría)
- Respuestas según una escala de Likert con **7 opciones**

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

Usefulness

- It helps me be more effective.
- It helps me be more productive.
- It is useful.
- It gives me more control over the activities in my life.
- It makes the things I want to accomplish easier to get done.
- It saves me time when I use it.
- *It meets my needs.*
- It does everything I would expect it to do.

Ease of Use

- It is easy to use.
- It is simple to use.
- It is user friendly.
- It requires the fewest steps possible to accomplish what I want to do with it.
- *It is flexible.*
- *Using it is effortless.*
- *I can use it without written instructions.*
- *I don't notice any inconsistencies as I use it.*
- *Both occasional and regular users would like it.*
- *I can recover from mistakes quickly and easily.*
- *I can use it successfully every time.*

Ease of Learning

- I learned to use it quickly.
- I easily remember how to use it.
- It is easy to learn to use it.
- *I quickly became skillful with it.*

Satisfaction

- I am satisfied with it.
- I would recommend it to a friend.
- It is fun to use.
- It works the way I want it to work.
- It is wonderful.
- I feel I need to have it.
- It is pleasant to use.

Users rate agreement with these statements on a seven-point Likert scale, ranging from strongly disagree to strongly agree. Statements in *italics* were found to weight less heavily than the others.

Fig. a

The USE questionnaire. From the work of Lund (2001); used with permission.

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

- Para algunos cuestionarios, como USE o CSUQ, los resultados se representan en un **gráfico radar** (ejemplo para cuestionario USE: *Fig. a*)
 - La comparación de gráficos permite comparar formas geométricas similares para reconocer patrones de usabilidad

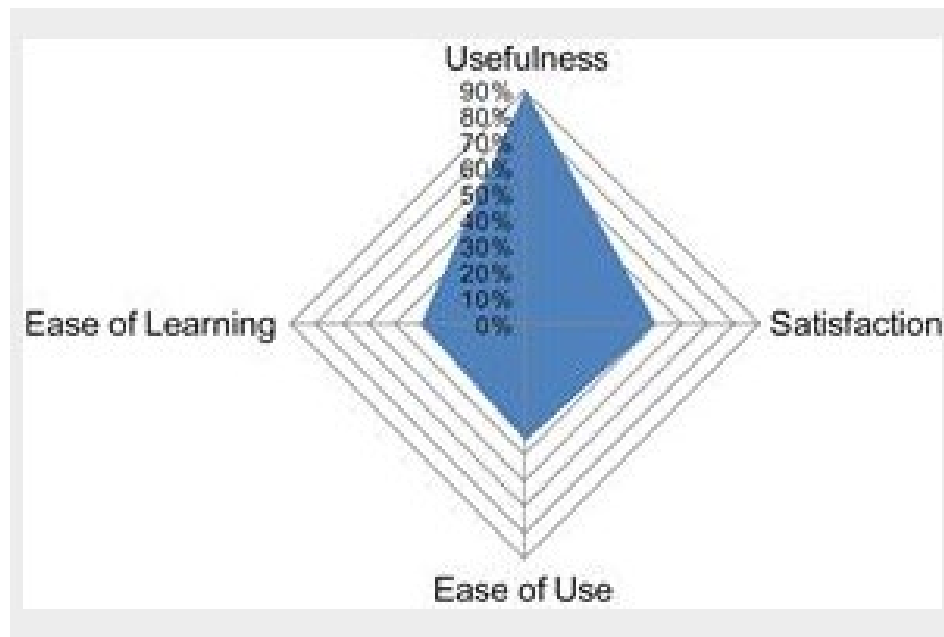


Fig. a

7.3 Experimentación con Usuarios.

Métricas subjetivas. Evaluación post-sesión

- **Comparación** de métricas subjetivas post-sesión
 - En un estudio de Tullis y Stetson, 2004, se compararon {*SUS*, *QUIS*, *CSUQ*}, concluyendo que **SUS es más exacto**
 - Posibles causas: 1) *SUS* utiliza afirmaciones positivas y negativas, estimulando una mayor atención de los participantes; 2) No hace agrupaciones de preguntas relativas a diferentes categorías
 - *SUS* se ha usado para hacer diversas comparaciones, por ejemplo:
 - Windows-ME vs Windows-XP, American Institutes for Research, 2001 (56% vs 74%, $p\text{-value} < 0.0001$)

7.3 Experimentación con Usuarios.

Métricas subjetivas web

- Aplicables en una evaluación web post-tarea o post-sesión
- Técnicas
 - **1) WAMMI (*Website Analysis and Measurement Inventory*)**
 - **2) ACSI (*American Customer Satisfaction Index*)**
- **1) WAMMI**
 - WAMMi está formado por **20 afirmaciones**, algunas positivas y otras negativas, con respuestas según una escala de **5 opciones**
 - Ha sido usado en cientos de sitios web, creando un resultado para un sitio web nuevo comparando con los resultados ya existentes.
 - El resultado contiene **una medida de usabilidad global**
 - Sus **BD de resultados sirven de punto de referencia** (característica común a muchas variantes de cuestionarios -- *benchmarking*)

7.3 Experimentación con Usuarios.

Métricas subjetivas web

- **2) ACSI**

- Ha sido aplicado a cientos de sitios web de administración, Amazon, ...
- Está formado por **14 preguntas** con respuestas en **escala de 10 puntos**
- Los resultados proporcionan
 - **Valores para 6 categorías:** contenido, funcionalidad, apariencia, navegación, búsqueda y rendimiento
 - Un resultado de **satisfacción global**
 - Valores de **comportamiento probable futuro** para “probabilidad de volver” y “recomendaría a otros”

7.3 Experimentación con Usuarios.

Datos considerados en usabilidad web

- 1) **Datos habituales** (*web analytics*)
 - Visitantes, visitas, visitas a página, ... (*Google Analytics* , ...)
- 2) Tasas de **seguimiento de vínculos** (*click-through*)
 - Uso de un determinado enlace o botón para acceder a otra página
- 3) **Tasas de abandono** (*drop-off*)
 - Detectan contextos con importantes problemas de usabilidad
- 4) **Datos resultados de pruebas A/B**
 - Se muestran aleatoriamente **páginas en pruebas** a usuarios para detectar problemas de usabilidad (estas pruebas son frecuentes en Amazon, Facebook , Google, etc.).

7.3 Experimentación con Usuarios.

Métricas psicológicas y de comportamiento

• 1) Expresiones de usuarios

- ¡perfecto!, ¡esto es horrible!, ...

• 2) Rastreo ocular (*eye tracking*)

- Ejemplo: patrones con forma de F en páginas web ([URL](#)), *Fig. a*

• 3) Detectores de emociones

- Expresiones de cara
- Conductancia de piel
- Electroencefalogramas

• 4) Otros aspectos

- Ritmo cardíaco, *Fig. b*
- Presión sobre el ratón
- Presión sobre asiento

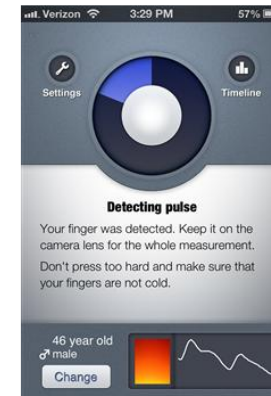


Fig. b

Example of Azumio Stress Checker app for the iPhone that measures stress through HRV by detecting heart rate through the camera.



Heatmaps from user eyetracking studies of three websites. The areas where users looked the most are colored red, the yellow areas indicate fewer views, followed by the least-viewed blue areas. Gray areas didn't attract any fixations.

Fig. a

7.3 Experimentación con Usuarios.

Métodos de evaluación en UX

- **Existen muchos métodos de evaluación para UX**
 - En general, permiten medir satisfacción subjetiva o percepción del usuario ([URL1](#), [URL2](#))
 - Clasificación según criterios:
 - *Método de evaluación*
 - Campo/Laboratorio/Online/Cuestionarios
 - *Fase de desarrollo*
 - Conceptual/PrimerosPrototipos/PrototiposFuncionales/ProductosDeMercado
 - *Periodo de uso*
 - PrevioUso/DuranteInteraccion/DuranteTarea/LargoPlazo
 - *Tipos de evaluadores*
 - ExpertosUX/UnicoUsuario/GruposDeUsuarios/ParesDeUsuarios

7.3 Experimentación con Usuarios.

Combinación de métricas

- Dos modos de combinación de métricas (suponemos la **ejecución de varias tareas** por cada usuario):
 - **1) Combinación de varias métricas de usuario**
 - A) Basada en un **objetivo combinado** (p.ej. *Tab. a*)
 - B) Basada en **porcentajes** (p.ej. *Tab. b*)
 - Se asocian **pesos** a las métricas, expresando éstas en porcentajes. Se pueden **combinar métricas de rendimiento y subjetivas**.
 - **2) Comparación de datos con resultados ideales, o de expertos** (p.ej: *Fig. a*)

7.3 Experimentación con Usuarios.

Combinación de métricas

Tab. a

Sample task completion and task time data from eight participants^a.

Participant #	Task Completion	Task Time (secs)	Goal Met?
1	85%	68	1
2	70%	59	0
3	80%	79	0
4	75%	62	0
5	90%	72	0
6	80%	60	1
7	80%	56	1
8	95%	78	0
Average:	82%	67	38%

^aAlso shown are averages for task completion and time and an indication of whether each participant met the objective of completing at least 80% of the tasks in no more than 70 seconds.

Objetivo combinado → completar 80% de tareas en menos de 70 segundos

7.3 Experimentación con Usuarios.

Combinación de métricas

Tab. b

Calculation of weighted averages^a.

Participant #	Time	Weight	Tasks	Weight	Rating	Weight	Weighted Average
1	38%	1	47%	1	60%	2	51%
2	50%	1	60%	1	65%	2	60%
3	74%	1	87%	1	78%	2	79%
4	36%	1	40%	1	43%	2	40%
5	89%	1	73%	1	80%	2	81%
6	48%	1	60%	1	83%	2	68%
7	43%	1	53%	1	63%	2	55%
8	42%	1	47%	1	35%	2	40%
9	100%	1	60%	1	95%	2	88%
10	45%	1	67%	1	90%	2	73%

Average → 63%

^aEach individual percentage is multiplied by its associated weight, these products are summed, and that sum is divided by the sum of the weights (4, in this example).

Métricas expresadas en % y con peso: tiempo, tareas-completadas, valoración

7.3 Experimentación con Usuarios.

Combinación de métricas

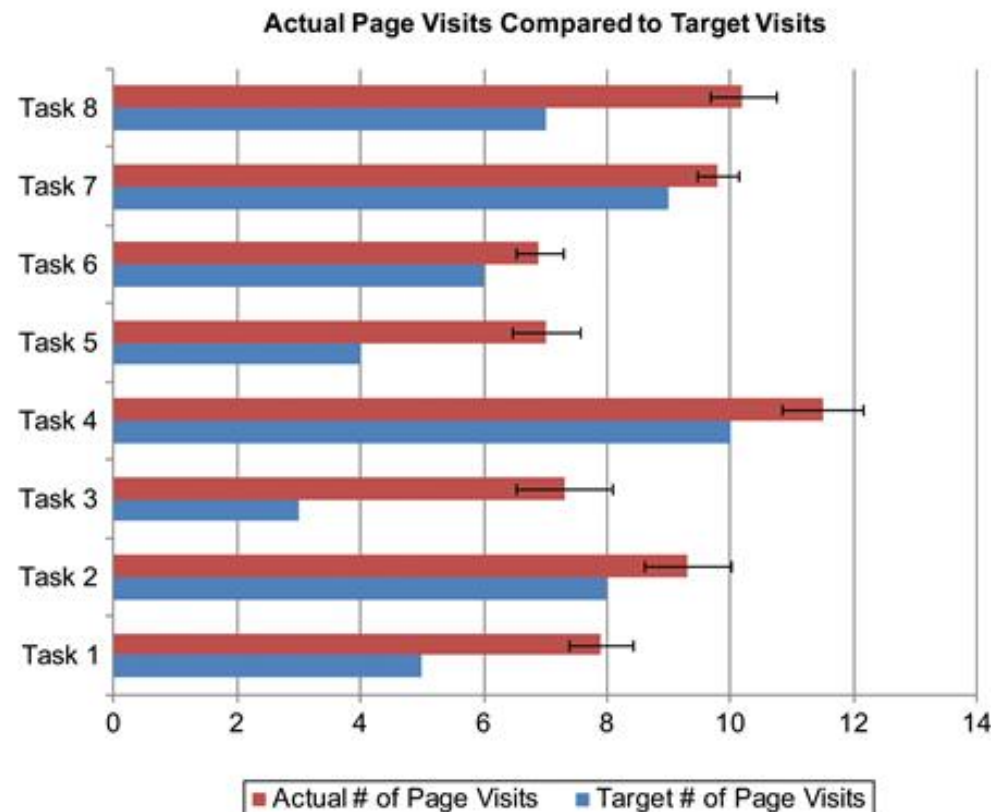


Fig. a

Target and actual number of page visits for each of eight tasks. Error bars represent the 90% confidence interval for the actual number of page visits.

7.3 Experimentación con Usuarios.

Métricas en ordenación de tarjetas

- La ordenación de tarjetas es una técnica para generación de **agrupaciones de datos**
 - Aplicada habitualmente en **contextos web** (cada página es una carta)
- Existen 2 variantes
 - **A) Ordenación abierta.** Cada usuario crea una agrupación de los tarjetas y puede crear un número arbitrario de grupos.
 - **B) Ordenación cerrada.** Se parte de un conjunto cerrado de grupos, cada grupo con un nombre. A continuación cada usuario asocia cada tarjeta a un grupo.
- Tras realizar ordenaciones abiertas o cerradas sobre un conjunto de usuarios, existen diversos métodos para generar **una agrupación resultado** de las tarjetas
 - Además, existen **métricas para estimar la idoneidad** (*fitness*) de la agrupación resultado
- **El número ideal de participantes** es 10-15 (ha sido comprobado en algunos estudios de ordenación abierta)

7.3 Experimentación con Usuarios.

Métricas en ordenación de tarjetas

- **A) Ordenación abierta**

- Punto de partida:
 - Matriz de distancias para cada participante, p.ej: *Tab. A*
 - Condición → la relación binaria “estar en el mismo grupo” debe ser de equivalencia (reflexiva, simétrica y transitiva)
 - Matriz global de distancias (varios participantes), p.ej: *Tab. b* para 20 participantes
 - Número de agrupaciones posibles (*Bell numbers*, [URL](#))
 - 1, 2, 5, 15, 52, 203, 877, 4.140, 21.147, 115.975, ...
- Métodos estadísticos para generar una agrupación resultado
 - **1) HCA (*Hierarchical Cluster Analysis*)**
 - **2) MDS (*Multidimensional Scaling*)**

7.3 Experimentación con Usuarios.

Métricas en ordenación de tarjetas

- **1) HCA (*Hierarchical Cluster Analysis*)**
 - Colocación en ramas de las tarjetas. Las tarjetas que se unen antes en el gráfico son más similares que las que lo hacen después.
 - Existen diferentes algoritmos para crear los enlaces a partir de la matriz global de distancias
 - Ejemplo: *Fig. a* es el resultado de HCA para *Tab. b*
 - Algoritmo de creación de enlaces (recursivo; coger el par más cercano, crear su rama, crear una tarjeta para ese grupo considerando en matriz de distancias el promedio, ...)
 - La línea vertical crea una agrupación con 4 grupos (intersección con líneas horizontales)
 - ¿Dónde lanzar la línea vertical?. Una posible aproximación es buscar una agrupación resultado cuyo número de grupos sea el promedio de grupos creados por los participantes.
 - Métrica de idoneidad
 - Promedio de idoneidades para todos los participantes. Por ejemplo, en el participante de *Tab. a*, de entre los 45 posibles pares, hay 7 que no coinciden (apple-tomato, ...), lo cual supone un 84% de idoneidad.

7.3 Experimentación con Usuarios.

Métricas en ordenación de tarjetas

Tab. a

Distance matrix for one participant in the fruit card-sorting example.

[illegible]

Tab. b

Overall distance matrix for 20 participants in the fruit card-sorting study.

[illegible]

7.3 Experimentación con Usuarios.

Métricas en ordenación de tarjetas

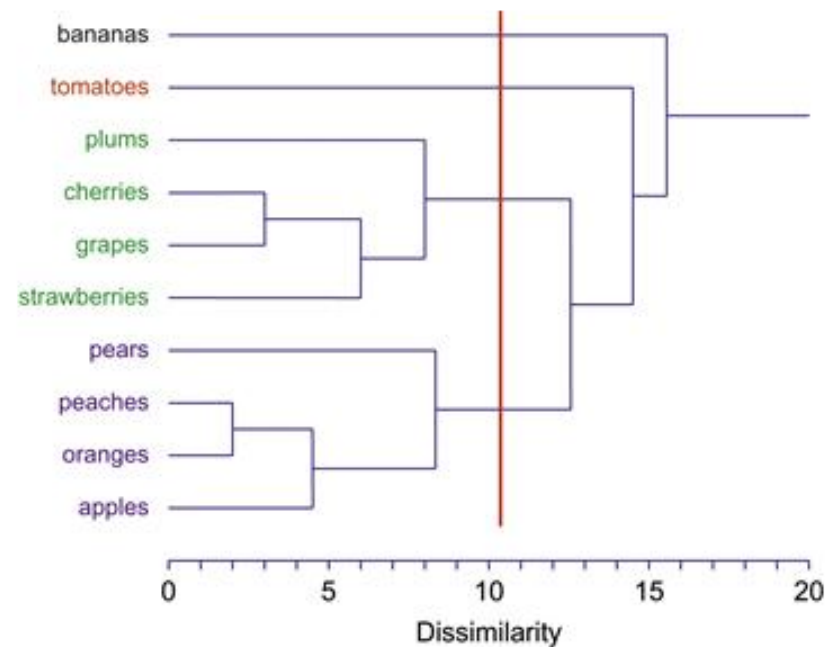


Fig. a

Result of a hierarchical cluster analysis of data shown in *Tab. b*

7.3 Experimentación con Usuarios.

Métricas en ordenación de tarjetas

- **2) MDS (*Multidimensional Scaling*)**

- A partir de la distancia entre pares de la matriz global de distancias (p.ej: *Tab. b*) el algoritmo encuentra una representación N-dimensional de las tarjetas de modo que las distancias entre pares sean similares a las de la matriz global, visualizando la agrupación resultado en N dimensiones
 - Es como si se creara un mapa a partir de distancias entre ciudades, si $N=2$. El N puede ser arbitrario, pero obtienen resultados significativos para $N=\{2,3\}$

- **B) Ordenación cerrada**

- Menos habitual que una ordenación abierta
- Se suele comenzar con una o varias ordenaciones abiertas para obtener varias agrupaciones resultado. A continuación, se realizan ordenaciones cerradas para seleccionar la mejor de entre las agrupaciones candidatas
 - Se define una métrica de idoneidad para cada ordenación cerrada

Bibliografía y Referencias Específicas – Tema 7

- Thomas Tullis, William Albert, "***Measuring the User Experience***", 2nd Edition, Morgan Kaufmann, 2013, [URL \(safari\)](#)
- Jonathan Lazar; Jinjuan Heidi Feng; Harry Hochheiser, "***Research Methods in Human-Computer Interaction***", 2nd Edition, Morgan Kaufmann, 2017, [URL \(safari\)](#), [companion website](#)
- Rex Hartson, Pardha S. Pyla, "*The UX Book*", Morgan Kaufmann, 2012, [URL \(safari\)](#)
- Jeffrey Rubin, Dana Chisnell, Jared Spool, "*Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*", 2008, [URL \(safari\)](#), [companion website](#)
- MIT OpenCourseWare, "*User Interface Design and Implementation*" , 2011, [URL](#)
- Granollers, T., Lorés, J. Cañas, J. "*Diseño de Sistemas Interactivos Centrados en el Usuario*". Editorial UOC, 2005
- Julio Abascal, José J. Cañas, Miguel Gea, Ana Belén Gil, Jesús Lorés, Ana Belén Martínez Prieto, Manuel Ortega, Pedro Valero, Manuel Vélez, "*La Interacción Persona-Ordenador*", 2001, libro AIPO
- Stanford Coursera, "*Human-Computer Interaction*", 2019, [URL](#)