

# Trabajo 1: Diseño de un experimento controlado

*Daniel Jiménez García*  
*Sergio Muñoz Gómez*  
*Ángela Caiqing Pousada Morán*

*Fecha: 8 de octubre de 2025*

# Resumen

Añadir resumen después de terminar el trabajo

# Capítulo 1

## Motivación

### 1.1. Problema a investigar

En la actualidad existen numerosos métodos para evaluar la usabilidad y experiencia de usuario (UX) a la hora de utilizar sistemas interactivos. Sin embargo, no existe un consenso claro sobre cuál de ellos resulta más adecuado para cada sistema. El problema que se aborda en este trabajo es la falta de evidencia empírica que permita comparar la **efectividad, eficiencia y satisfacción de los evaluadores** cuando aplican un método empírico (*A/B Testing*) frente a un método de inspección grupal (*Recorrido Pluralista*).

### 1.2. Definición del experimento

Analizar los métodos de evaluación de usabilidad **A/B Testing** y **Recorrido Pluralista** con el propósito de comparar su capacidad para detectar problemas de usabilidad y medir su impacto en la satisfacción y eficiencia de los evaluadores.

### 1.3. Contexto

El experimento se llevará a cabo en un entorno controlado de laboratorio, sobre un prototipo web de complejidad media. Se contará con estudiantes de ingeniería de software y diseño UX como participantes. Cada sesión incluirá la realización de tareas específicas dentro del sistema y la evaluación mediante uno de los dos métodos propuestos.

## Capítulo 2

### Trabajos relacionados

Diversos estudios previos han comparado métodos de evaluación de usabilidad. Nielsen (1994) definió la *Evaluación Heurística* como técnica de inspección rápida, mientras que Rubin y Chisnell (2008) destacaron la importancia de los métodos empíricos como el *User Testing* para observar comportamientos reales.

El **A/B Testing** se utiliza ampliamente en el ámbito web para medir diferencias en satisfacción y eficiencia entre dos versiones de una interfaz (Kohavi et al., 2009). Por su parte, el **Recorrido Pluralista** (Bias, 1994) combina la revisión de expertos, diseñadores y usuarios, permitiendo identificar problemas desde perspectivas diversas.

A pesar de su uso extendido, no existen suficientes estudios comparativos entre ambos métodos bajo condiciones controladas, especialmente considerando variables subjetivas como satisfacción o percepción de facilidad de uso.

# Capítulo 3

## Descripción del diseño

### 3.1. Hipótesis y variables

#### Hipótesis:

- $H_{01}$ : No existen diferencias significativas en la **satisfacción** entre A/B Testing y Recorrido Pluralista.
- $H_{11}$ : A/B Testing genera mayor satisfacción percibida que el Recorrido Pluralista.
- $H_{02}$ : No existen diferencias significativas en la **usabilidad percibida** entre los métodos.
- $H_{12}$ : El Recorrido Pluralista detecta más problemas de usabilidad que A/B Testing.
- $H_{03}$ : No existen diferencias significativas en la **eficiencia** (tiempo medio de tareas).
- $H_{13}$ : Los evaluadores son más eficientes (menor tiempo) usando A/B Testing.

#### Variables:

- **Variable independiente:** Método de evaluación (A/B Testing, Recorrido Pluralista).
- **Variables dependientes:**
  - **Satisfacción:** medida mediante cuestionario Likert (1–7).
  - **Usabilidad:** puntuación SUS (*System Usability Scale*).
  - **Eficiencia:** tiempo promedio (en segundos) para completar tareas.
- **Variables de control:** tipo de tarea, complejidad del prototipo, experiencia previa de los participantes.

### 3.2. Diseño del experimento

El experimento seguirá un diseño **intra-sujetos contrabalanceado**. Cada participante aplicará ambos métodos (A/B Testing y Recorrido Pluralista) sobre dos conjuntos de tareas diferentes, en sesiones separadas por 24 horas para mitigar los efectos de aprendizaje y fatiga. El orden de aplicación de los métodos será aleatorizado para contrarrestar sesgos de orden.

### 3.3. Selección de sujetos

Se seleccionarán entre 16 y 20 participantes, con formación básica en usabilidad o experiencia de usuario. Los criterios de inclusión son:

- Familiaridad con interfaces web.
- No haber participado previamente en experimentos similares.

Cada participante firmará un consentimiento informado y completará un cuestionario demográfico inicial.

### 3.4. Objetos e instrumentación

Los objetos experimentales serán dos versiones de una aplicación web (versión A y versión B) con diferencias en diseño o disposición visual.

- **Instrumentos de medida:** - Cronómetro o software de registro (tiempo por tarea). - Cuestionario SUS (Brooke, 1996). - Escala de satisfacción Likert (1–7). - Registro de observaciones y errores detectados.

### 3.5. Evaluación de la validez

**Validez interna:** se controlará el efecto orden mediante contrabalanceo. Se mantendrá la misma complejidad de tareas y condiciones experimentales.

**Validez externa:** los resultados serán aplicables a contextos similares (evaluaciones de usabilidad de aplicaciones web con participantes semiexpertos).

**Validez de constructo:** las métricas SUS y Likert se seleccionan por su fiabilidad y validez en estudios de usabilidad.

**Validez de conclusión:** se emplearán pruebas t pareadas (o Wilcoxon si no hay normalidad) con  $\alpha = 0,05$ , y tamaños del efecto (Cohen's d) para cuantificar diferencias.

# Referencias

- Esto es referncia buscada sobre qué es el pluralistic walk...[1]

Lo de abajo es puro ChatGPT(no me escondo)

- Bias, R. (1994). *The Pluralistic Usability Walkthrough: Coordinated Empathies*. In Nielsen & Mack (Eds.), Usability Inspection Methods.
- Brooke, J. (1996). *SUS: A Quick and Dirty Usability Scale*. Usability Evaluation in Industry.
- Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. (2009). *Controlled experiments on the web: Survey and practical guide*. Data Mining and Knowledge Discovery, 18(1).
- Nielsen, J. (1994). *Usability Engineering*. Morgan Kaufmann.
- Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley.
- Wohlin, C., et al. (2012). *Experimentation in Software Engineering*. Springer.

# Bibliografía

- [1] Chauncey Wilson. «Chapter 5 - Pluralistic Usability Walkthrough». En: *User Interface Inspection Methods*. Ed. por Chauncey Wilson. Boston: Morgan Kaufmann, 2014, págs. 81-97. ISBN: 978-0-12-410391-7. DOI: <https://doi.org/10.1016/B978-0-12-410391-7.00005-1>. URL: <https://www.sciencedirect.com/science/article/pii/B9780124103917000051>.



# Apéndice A

## Anexo I: Tareas de evaluación

Cada participante completará un conjunto de tareas representativas en la aplicación web.

- **Tarea 1:** Localizar un producto y completar una compra simulada.
- **Tarea 2:** Cambiar una configuración de usuario en el perfil.

En A/B Testing, se compararán tiempos y errores entre versiones A y B. En el Recorrido Pluralista, se discutirán los pasos de interacción con un grupo de tres evaluadores (usuario, diseñador y experto).

## Apéndice B

### Anexo II: Cuestionario de satisfacción y usabilidad

- Escala SUS (10 ítems, 1–5).
- Escala de satisfacción general (Likert 1–7).
- Preguntas abiertas sobre percepción del método aplicado.