MATERIAL DE LECTURA OBLIGATORIA / UNIDAD 05

Validación de Interfaces y Evaluación de la Experiencia de Usuario

Del prototipo navegable a las decisiones basadas en evidencia



Observar · Interpretar · Mejorar | El diseño no termina cuando el prototipo se ve terminado.

Asignatura
Técnicas de UX y Diseño de Interfaces
Ingeniería de Sistemas · UNICEN

Enfoque
Métodos de inspección, pruebas con usuarios, métricas e iteración del Proyecto Integrador

Material académico de apoyo · Noviembre de 2026 · Lectura obligatoria · Edición ampliada

ORIENTACIÓN DE LECTURA

Contenido de la unidad

Del análisis experto a la evaluación con personas y la mejora trazable del diseño.

Este apunte amplía los fundamentos y desarrolla procedimientos aplicables a un prototipo. La lectura propone un recorrido continuo: primero comprender qué se evalúa, luego elegir un método, realizar la prueba y finalmente documentar decisiones. Los ejemplos de canje son ficticios; las condiciones del TP se indican en la sección de integración.

01 De diseñar a validar
02 Evidencia, hipótesis y tipos de evaluación
03 Fidelidad: elegir qué vale la pena probar
04 Inspección y pruebas con usuarios
05 Evaluación heurística: fundamentos y procedimiento
06 Las diez heurísticas en un flujo real
07 Escala de severidad: no todos los problemas pesan igual
08 Ficha de hallazgo: un ejemplo completo
09 Recorrido cognitivo: evaluar cómo se aprende
10 Un recorrido cognitivo resuelto
11 Diseñar el estudio antes de reclutar
12 Escenarios, tareas y criterios de éxito
13 Think Aloud: escuchar mientras la persona actúa
14 Roles, ubicación y neutralidad
15 Cuándo intervenir y cómo formular preguntas
16 Consentimiento, privacidad y contexto de prueba
17 Pruebas moderadas, no moderadas y remotas
18 Una sesión de 10–15 minutos, de principio a fin
19 Observar, registrar e interpretar
20 Cinco usuarios: alcance y límites de la evidencia
21 Eficacia, eficiencia y errores
22 Ejemplo resuelto: cinco participantes
23 SEQ: facilidad percibida por tarea
24 SUS: evaluación global de la usabilidad percibida
25 Qué representan y qué no representan
26 Evaluación predictiva y asistencia de inteligencia artificial
27 Del registro individual a un hallazgo
28 Decidir qué corregir primero
29 Validar hipótesis mediante evidencia
30 El caso completo y su relación con el TP 5
31 Instrumento de planificación y registro
32 Preguntas para revisar lo aprendido

> Objetivo: al finalizar la unidad, el estudiante será capaz de planificar una evaluación, observar tareas sin inducir respuestas, interpretar métricas y justificar modificaciones del diseño con evidencia.

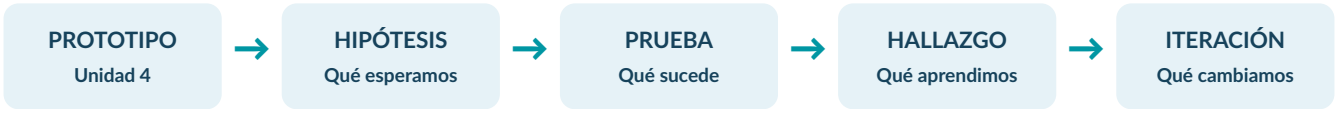
01 / FUNDAMENTOS

De diseñar a validar

Evaluar es producir evidencia sobre la calidad de una interacción para decidir qué conservar, qué modificar y qué investigar todavía.

Una interfaz puede cumplir su especificación técnica y aun así impedir que una persona complete su objetivo. Un botón puede responder correctamente a un clic, por ejemplo, aunque su etiqueta no permita reconocer qué sucederá al pulsarlo. Por eso las pruebas funcionales de software y la evaluación de experiencia de usuario se complementan: unas verifican el funcionamiento previsto, las otras examinan la relación entre el sistema, la persona y su tarea.

La ISO 9241-11:2018 sitúa la usabilidad en un contexto de uso especificado. No es una cualidad absoluta de una pantalla: depende de quién la usa, con qué propósito, en qué entorno y con qué recursos. Conviene observar tres dimensiones: **eficacia** (alcanzar el objetivo correctamente), **eficiencia** (recursos empleados para hacerlo) y **satisfacción** (experiencia percibida). Una aplicación rápida puede ser ineficaz si resuelve la operación equivocada; una tarea exitosa puede resultar agotadora.



Del diseño centrado en el usuario a la evaluación

En la Unidad 1 se investigó el problema; en la 2 se estudiaron principios de interacción; en la 3 se organizaron los recorridos; en la 4 se desarrolló el prototipo. Ahora se vuelve a mirar el diseño desde el objetivo de una persona. Evaluar temprano reduce el costo de modificar decisiones antes de que se conviertan en código y dependencias. También se evalúa después de publicar, porque el entorno real puede revelar problemas que una maqueta no reproduce.

Dimensión	Pregunta	Evidencia en un canje
Eficacia	¿Consiguió el beneficio correcto?	Confirmación correcta sin ayuda.
Eficiencia	¿Qué esfuerzo necesitó?	Tiempo, retrocesos y errores.
Satisfacción	¿Cómo vivió la operación?	Facilidad percibida y comentarios.

> Evaluar no es pedir una aprobación estética. Es determinar si el diseño permite actuar, comprender los resultados y recuperarse de problemas.

Referencias: [1], [22] (ver bibliografía).

02 / FUNDAMENTOS

Evidencia, hipótesis y tipos de evaluación

Una evaluación útil empieza por una pregunta que pueda responderse con un procedimiento explícito.

Una **hipótesis de diseño** es una expectativa contrastable sobre la relación entre una característica de la interfaz y un comportamiento. Por ejemplo, si el saldo de puntos permanece visible durante el canje, se espera que las personas puedan reconocer qué beneficios están a su alcance sin salir de la pantalla. No basta afirmar que «la pantalla es intuitiva»: hay que definir qué conducta demostraría que la expectativa se cumple.

Una **evaluación formativa** busca descubrir problemas y orientar cambios durante el desarrollo. Es el enfoque central del curso. Una evaluación sumativa, en cambio, compara resultados con objetivos definidos, requisitos o una referencia de desempeño. Puede realizarse con un producto terminado o suficientemente estable y exige condiciones controladas y una muestra acorde a la conclusión pretendida.

Tres fuentes que responden preguntas diferentes

Fuente	Qué permite conocer	Ejemplo
Inspección	Riesgos a partir de principios y conocimiento experto.	El botón no refleja el estado pendiente.
Conducta	Acciones observadas en tareas concretas.	P02 vuelve al catálogo después de confirmar.
Percepción	Valoración de la persona sobre la experiencia.	P02 declara que no sabe si terminó.

La combinación de evidencias, conocida como **triangulación**, ayuda a distinguir causas posibles. Si una persona dice que la operación fue fácil pero cometió un error, no hay que elegir entre su opinión y el dato: ambos describen aspectos diferentes. Quizá consideró la tarea familiar, aunque no advirtió que seleccionó una opción equivocada. La evaluación debe registrar esa contradicción en lugar de borrarla.

Qué cambiaría una decisión

Antes de convocar participantes, escriban qué resultado los llevaría a revisar el diseño. Si el objetivo es descubrir incertidumbre tras confirmar, observar repeticiones del botón o dudas verbalizadas orientará la revisión del feedback. No se trata de fijar un porcentaje arbitrario con pocos usuarios, sino de establecer qué constituye evidencia útil y evitar interpretar el estudio según lo que el equipo desea encontrar.

> Un principio orienta una hipótesis; una observación aporta evidencia. Ninguno basta, por separado, para demostrar que una solución es adecuada.

Referencias: [1], [4], [22] (ver bibliografía).

03 / PROTOTIPOS

Fidelidad: elegir qué vale la pena probar

Un prototipo es una representación parcial del producto que permite aprender sin construir toda la solución.

La fidelidad indica cuánto aproxima una representación a la experiencia prevista. Puede variar en lo visual, en el contenido y en el comportamiento. Un boceto puede simular una navegación completa con hojas de papel; una maqueta de gran detalle visual puede carecer de vínculos entre pantallas. No existe una correspondencia automática entre calidad gráfica y capacidad de evaluación.

Nivel	Qué contiene	Pregunta que responde
Baja	Bocetos, wireflows y textos provisionales.	¿Tiene sentido este recorrido? ¿Falta un paso?
Media	Wireframes navegables, etiquetas y estructura más precisas.	¿Se encuentra la opción? ¿Se entiende el orden?
Alta	Pantallas con identidad visual, datos plausibles y estados simulado-realistas.	¿Se comprende la jerarquía y se completa la tarea?

Ejemplo: una reserva de turno

Con papel se puede explorar si la persona espera elegir primero el profesional o la fecha. En fidelidad media, se prueban calendarios, rótulos y alternativas de horario. En alta fidelidad, se observa si el horario elegido se reconoce, si el botón principal se distingue y si la confirmación resulta comprensible. Ninguno de esos prototipos verifica por sí mismo la disponibilidad real de un turno ni la integridad de una base de datos.

Alcance y preparación

El nivel adecuado depende de la pregunta y del costo de modificar la solución. Si se quiere investigar la arquitectura, prototipar animaciones complejas resulta innecesario. Si se quiere evaluar una interacción que exige corregir datos o volver atrás, esa posibilidad debe estar construida o simulada de forma consistente. Las personas no deben atribuirse errores provocados por pantallas que el prototipo no representa.

> Para el TP 5 se utiliza el flujo navegable de la Unidad 4. Verifiquen que los caminos necesarios funcionen y que la operación termine en un estado reconocible, con datos ficticios y sin transacciones reales.

Referencias: [2], [4] (ver bibliografía).

04 / MÉTODOS

Inspección y pruebas con usuarios

La inspección analiza el diseño mediante criterios. La prueba de usabilidad estudia cómo las personas actúan frente a ese diseño.

En una evaluación por inspección, los evaluadores recorren una interfaz y la contrastan con principios de interacción. No necesitan convocar a usuarios para identificar un estado de error ausente o una operación irreversible sin confirmación. El resultado consiste en una lista de riesgos razonados. Sin embargo, los evaluadores conocen el producto o tienen experiencia técnica y pueden pasar por alto obstáculos que encuentran quienes no comparten ese conocimiento.

En la prueba de usabilidad, se presenta una tarea a personas próximas al público objetivo y se observa si logran resolverla y cómo. La prueba permite descubrir estrategias no previstas por los diseñadores. No equivale a una entrevista de descubrimiento: en la entrevista se reconstruye una experiencia relatada; aquí el foco es una experiencia que sucede en el momento y puede observarse.

Método	Unidad de análisis	Resultado
Evaluación heurística	Pantallas, estados y decisiones de interacción.	Riesgos agrupados por principios.
Recorrido cognitivo	Pasos de una tarea para un usuario nuevo.	Obstáculos de aprendibilidad.
Prueba moderada	Personas realizando tareas bajo observación.	Conductas, errores, expectativas y métricas.
Prueba no moderada	Intentos registrados con un protocolo autónomo.	Resultados comparables bajo condiciones definidas.

Cómo combinarlos

Primero inspeccionen el prototipo para evitar que problemas evidentes consuman la sesión; corrijan los bloqueos que impiden probar la tarea. Después observen usuarios reales para saber si los riesgos restantes se manifiestan y descubrir otros. Finalmente, documenten qué problemas provienen de una inspección y cuáles están respaldados por sesiones. Un riesgo que todavía no se observó no desaparece: puede requerir otra tarea, otro perfil o una implementación más completa.

> No es necesario elegir un único método para todo el proyecto. El método se selecciona según la pregunta, el estado del producto y la evidencia disponible.

Referencias: [1], [3], [4], [6] (ver bibliografía).

05 / INSPECCIÓN

Evaluación heurística: fundamentos y procedimiento

Las heurísticas son principios generales que permiten revisar decisiones de interfaz sin convertir las preferencias del evaluador en reglas.

Nielsen y Molich introdujeron la evaluación heurística como un método de inspección. Las heurísticas estudiadas en la Unidad 2 funcionan como lentes: ayudan a reconocer situaciones que pueden perjudicar la interacción, pero no son una lista de requisitos idéntica para todos los productos. Por ejemplo, la visibilidad del estado exige información oportuna sobre una operación, aunque su representación concreta dependa del contexto.

El procedimiento empieza delimitando el flujo, el perfil de usuario, el dispositivo y la versión del prototipo. Idealmente intervienen varios evaluadores, trabajando primero de forma independiente para reducir el arrastre por opiniones ajenas. Cada uno hace un recorrido general para entender la tarea y luego uno más detallado, revisando estados alternativos, errores y recuperación. Un único evaluador puede hacer el ejercicio, pero debe reconocer la limitación.



Registrar antes de discutir

Cada problema se registra por separado: lugar, condición que lo desencadena, evidencia visible, principio implicado y consecuencia prevista. Una misma dificultad puede relacionarse con varias heurísticas; elijan la que mejor explica el problema antes de agregar otras. No multipliquen hallazgos contando cada heurística como un problema distinto. Al terminar, consoliden coincidencias y asignen severidad sobre la lista unificada.

Ejemplo de una observación defendible

En la pantalla de confirmación, al tocar «Canjear», el control permanece habilitado y no aparece un cambio de estado. Esto incumple la visibilidad del estado del sistema y podría provocar reintentos. La ausencia de feedback se constata en la inspección; la frecuencia real de los reintentos sigue siendo desconocida.

> No escribir «es feo» ni «debería ser azul». Describir la condición observable y el efecto sobre una tarea permite discutir alternativas sin depender del gusto personal.

Referencias: [3], [5] (ver bibliografía).

06 / INSPECCIÓN

Las diez heurísticas en un flujo real

Repaso aplicado a una misma operación: de la lista teórica a preguntas concretas de inspección.

Para recorrer el prototipo, conviene tratar cada principio como una pregunta. La tabla propone ejemplos de problemas y de decisiones que podrían corregirlos. No todos aparecerán en un mismo producto; tampoco toda propuesta constituye una solución validada. La evaluación permite formular hipótesis que luego deben contrastarse.

Heurística	Qué revisar en un canje	Ejemplo de mejora
1. Visibilidad del estado	¿Se distingue procesamiento de resultado?	Mostrar estado pendiente y confirmación.
2. Mundo real	¿Se entienden puntos y condiciones?	Usar términos familiares y saldo explícito.
3. Control y libertad	¿Se puede volver sin perder la elección?	Permitir revisar antes de confirmar.
4. Consistencia	¿La misma acción se llama igual?	Usar «Canjear» consistentemente.
5. Prevención de errores	¿Es posible elegir algo inaccesible?	Informar saldo insuficiente antes de operar.
6. Reconocer antes que recordar	¿Hay que retener el saldo mentalmente?	Mostrar saldo junto a costo.
7. Flexibilidad y eficiencia	¿Usuarios frecuentes repiten pasos inútiles?	Acceso directo a beneficios recientes.
8. Simplicidad y relevancia	¿Los banners tapan la acción principal?	Jerarquizar el objetivo de la pantalla.
9. Recuperación de errores	¿El mensaje permite continuar?	Explicar causa y acción de recuperación.
10. Ayuda y documentación	¿Hay apoyo cuando se necesita?	Condiciones accesibles cerca de la decisión.

Al aplicar estas preguntas, revisen tanto el camino esperado como los estados de error. Una interfaz puede pasar la inspección en el caso feliz y fallar cuando hay saldo insuficiente, una red lenta o un dato inválido. Los componentes y estados trabajados en la Unidad 4 son parte del material que debe inspeccionarse.

> No se asigna severidad a una heurística, sino al problema concreto que perjudica la interacción.

Referencias: [3], [5] (ver bibliografía).

07 / INSPECCIÓN

Escala de severidad: no todos los problemas pesan igual

La severidad describe la importancia de corregir una dificultad de usabilidad; no es el costo de arreglarla.

La escala de Nielsen clasifica problemas de 0 a 4. Para decidir el nivel se considera su frecuencia, el impacto sobre quien lo encuentra y la persistencia de la dificultad. En una inspección la frecuencia suele estimarse; en una prueba se dispone de observaciones de la muestra. Es importante distinguir ambas fuentes en el informe.

Nivel	Interpretación	Ejemplo
0	No se identifica un problema.	El botón muestra claramente su estado.
1	Cosmético; no impide la tarea.	Alineación irregular en un texto auxiliar, sin efecto funcional.
2	Menor; causa fricción recuperable.	La etiqueta exige releer, pero se llega al beneficio.
3	Mayor; dificulta gravemente el objetivo.	La persona necesita ayuda para encontrar el canje.
4	Crítico; bloqueo esencial o riesgo grave.	Se canjea el beneficio incorrecto sin posibilidad de recuperación.

Cómo resolver desacuerdos de severidad

Si un evaluador asigna 2 y otro 4, no promedien mecánicamente. Pregunten en qué contexto está pensando cada uno: ¿el canje es reversible?, ¿hay devolución?, ¿qué sucede al repetir la operación? Una discrepancia puede revelar una regla de negocio que aún no se conoce. Documenten la gravedad provisional y la condición que permitiría resolverla. El nivel 4 debe reservarse para consecuencias críticas, no para todo detalle que desagrada al equipo.

Caso: botón de confirmación sin respuesta

El riesgo es mayor si exige ayuda o genera reintentos (3). Si la implementación permitiera un doble descuento irreversible, podría ser crítico (4). Sin comprobar el sistema real, no corresponde afirmar que el descuento duplicado ocurrió. La inspección sí debe dejar explícita esa posibilidad.

> Severidad y prioridad se relacionan, pero la prioridad también depende del riesgo operativo, la dependencia técnica y el esfuerzo de cambio.

Referencias: [5] (ver bibliografía).

08 / INSPECCIÓN

Ficha de hallazgo: un ejemplo completo

Un hallazgo de inspección debe ser reproducible, comprensible y útil para decidir una acción.

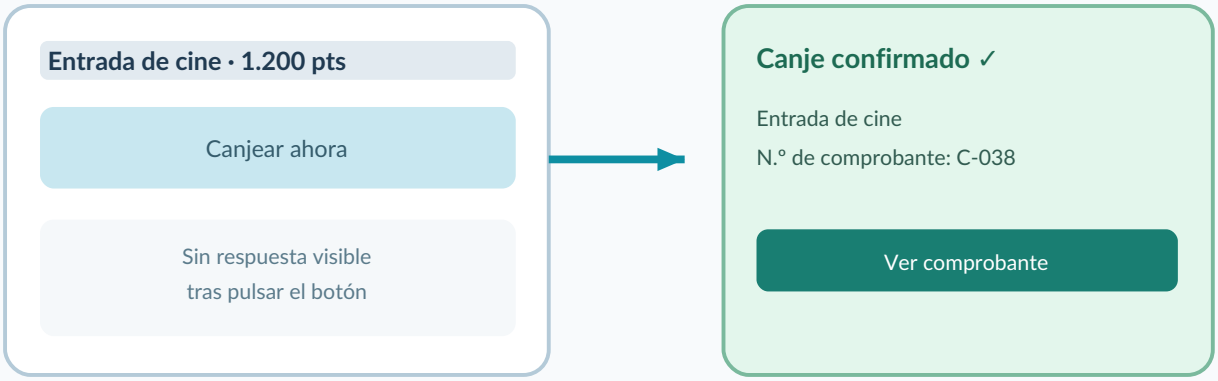


Ilustración didáctica: la ausencia de feedback puede interpretarse como operación fallida; la propuesta comunica estado y resultado.

Campo	Registro del evaluador
ID y origen	I-01 · Inspección · prototipo v1 · confirmación de canje
Condición	Después de pulsar «Canjear», la pantalla conserva el mismo contenido.
Evidencia	No cambia el texto del botón ni aparece un mensaje de confirmación.
Principio	Visibilidad del estado del sistema; prevención del error.
Efecto previsto	Duda sobre la finalización; posibilidad de repetir la acción.
Severidad	3 provisional; depende de si la operación puede duplicarse.
Cambio candidato	Estado pendiente, confirmación persistente y comprobante.
Verificación	Observar si el usuario reconoce el resultado sin ayuda ni reintentos.

La propuesta no se incorpora como si fuera una verdad definitiva. Puede que el usuario necesite además ver el nuevo saldo o que un comprobante no sea necesario para este servicio. El registro permite conservar el motivo del cambio y compararlo con la evidencia que se obtendrá en sesiones posteriores.

> La captura documenta el estado del diseño. La prueba con personas aporta evidencia de la dificultad que ese estado produce.

Referencias: [3], [5] (ver bibliografía).

09 / APRENDIBILIDAD

Recorrido cognitivo: evaluar cómo se aprende

El cognitive walkthrough estudia si una persona que no conoce el sistema puede descubrir cómo completar una tarea.

La técnica se basa en la idea de que las personas frecuentemente aprenden interfaces mediante exploración: formulan un objetivo, buscan una acción, la ejecutan e interpretan lo que ocurrió. El recorrido cognitivo simula esa secuencia desde el conocimiento que se supone que tiene un usuario nuevo. A diferencia de la evaluación heurística general, la atención se concentra en la **aprendibilidad** de un camino concreto.

Antes de realizarlo, definan quién es la persona, qué sabe del dominio, qué no conoce del producto y qué objetivo persigue. El equipo establece una secuencia correcta de acciones para la tarea (pueden existir varias rutas). Los evaluadores analizan cada paso, no solamente la pantalla inicial y final. Es conveniente que alguien registre las decisiones y los supuestos, mientras otra persona conduce la discusión.

Pregunta en cada paso	Qué se analiza
1. ¿Intentará producir el efecto adecuado?	Si la meta local coincide con lo que necesita hacer.
2. ¿Advertirá que la acción está disponible?	Visibilidad, ubicación y significantes del control.
3. ¿Relacionará la acción con su objetivo?	Comprensión del rótulo y consecuencias esperadas.
4. ¿Reconocerá el avance después de actuar?	Feedback y posibilidad de evaluar el nuevo estado.

Si alguna respuesta es negativa o incierta, redacten una historia de falla: qué podría intentar la persona, por qué y qué vería después. Eviten contestar «sí, porque sabemos que ahí está el botón». Precisamente el conocimiento del diseñador debe mantenerse separado del conocimiento del usuario previsto.

> El recorrido cognitivo es una inspección, no una sesión con usuarios. Las conclusiones son predicciones justificadas sobre el aprendizaje, no resultados de comportamiento observado.

Referencias: [6], [7], [21] (ver bibliografía).

10 / APRENDIBILIDAD

Un recorrido cognitivo resuelto

Ejemplo didáctico: persona que conoce los programas de beneficios, pero nunca utilizó esta aplicación.

La meta es obtener una entrada de cine con 1.800 puntos y comprobar que el canje quedó registrado. Se supone que la persona entiende qué son los puntos y ha utilizado una tienda web, pero desconoce las etiquetas particulares del sistema. La ruta prevista atraviesa el catálogo, el detalle y la confirmación.

Paso y acción correcta	Resultado de las cuatro preguntas	Riesgo detectado
1. Entrar al catálogo	Meta: sí. Visibilidad: dudosa. Relación: no, si se llama «Club». Feedback: sí, si abre listado.	El nombre «Club» no anticipa que allí se usan puntos.
2. Elegir una entrada	Meta: sí. Visibilidad: sí. Relación: sí, por título y costo. Feedback: sí, detalle abre.	Sin saldo visible, debe recordar cuántos puntos tiene.
3. Confirmar el canje	Meta: sí. Visibilidad: sí. Relación: dudosa si solo dice «Continuar». Feedback: no, pantalla no cambia.	No sabe si confirmó o si falta un paso.

Qué hacer con esas respuestas

El primer riesgo se vincula con etiquetado y modelo mental; el segundo con memoria de trabajo; el tercero con el golfo de evaluación estudiado en la Unidad 2. Se puede priorizar la confirmación porque afecta el resultado mismo de la operación. El catálogo podría probarse con una etiqueta más explícita y el detalle podría incluir el saldo disponible. No hace falta modificar todo antes de testear: elijan qué incógnitas requieren evidencia real.

Ejercicio para el proyecto

Seleccionen una tarea y escriban entre tres y cinco acciones previstas. Para cada acción, respondan las cuatro preguntas sin utilizar el conocimiento de cómo programaron la aplicación. Señalen al menos una duda razonable y redacten cómo la comprobarían con un participante.

> Una respuesta «no» indica un obstáculo posible. No permite afirmar que todas las personas fallarán en ese paso.

Referencias: [6], [7], [21] (ver bibliografía).

11 / PRUEBAS DE USABILIDAD

Diseñar el estudio antes de reclutar

Un estudio comienza con una decisión pendiente, no con una lista de pantallas que queremos mostrar.

Una prueba de usabilidad es una actividad de investigación en la que personas relevantes para el producto intentan realizar tareas mientras se registra su interacción. El diseño del estudio controla qué se presenta, qué se observa y cómo se interpreta. Una sesión improvisada puede aportar ideas, pero dificulta comparar usuarios si cada uno recibió instrucciones o datos diferentes.

Elementos de un plan de prueba

Elemento	Decisión que hay que documentar
Objetivo	Evaluar si se completa un canje sin instrucciones y se reconoce el resultado.
Participantes	Personas próximas al público identificado en la Unidad 1, ajenas al diseño.
Material	Prototipo v1, móvil o escritorio, saldo simulado de 1.800 puntos.
Tarea	Obtener entrada de cine de 1.200 puntos y reconocer la confirmación.
Criterio	Canje correcto y reconocimiento del resultado sin ayuda.
Registro	Éxito, tiempos, errores definidos, ayudas, comentarios y SEQ opcional.
Límites	Duración pactada, condiciones simuladas y problemas técnicos.

Antes de convocar, realicen una prueba piloto con alguien que ayude a detectar fallas del procedimiento. El piloto verifica que las instrucciones sean comprensibles, que la tarea sea posible y que el mecanismo de registro funcione. Si cambian el guion o el prototipo después del piloto, no mezclen esos datos con la ronda principal sin aclararlo.

> En el TP 5 basta una tarea transaccional completa. El alcance didáctico permite concentrarse en aprender a observar y analizar, en lugar de intentar evaluar toda la aplicación.

Referencias: [4], [8], [23] (ver bibliografía).

12 / PRUEBAS DE USABILIDAD

Escenarios, tareas y criterios de éxito

Una tarea realista describe la intención del usuario sin enseñar el camino que se busca evaluar.

El escenario proporciona el contexto necesario para actuar: quién es la persona, qué situación enfrenta y qué resultado necesita. La tarea concreta esa situación en una meta verificable. Debe ser comprensible sin utilizar palabras idénticas a las etiquetas que se quieren poner a prueba, salvo que esas palabras formen naturalmente parte del problema. El criterio de éxito es un acuerdo del equipo y no una instrucción que se lea al participante.

Evitar (enseña el camino)	Preferir (plantea una meta)
«Abrí Beneficios y tocá Canjear».	«Tenés 1.800 puntos y querés obtener una entrada de cine. Mostranos cómo lo harías».
«Andá al historial y descargá el PDF».	«Necesitás demostrar que la operación se realizó. Buscá algo que puedas mostrar».
«Usá el filtro de turnos de la izquierda».	«Necesitás un turno el viernes por la tarde. Intentá reservar uno».

Cómo decidir si la tarea se completó

Para el caso de canje, éxito independiente significa elegir el beneficio correcto, llegar al resultado esperado y reconocer que se registró. Éxito asistido significa llegar a ese resultado solo después de recibir una pista sustantiva. Un fracaso puede consistir en completar una operación equivocada, abandonar o alcanzar el límite de tiempo acordado. Un fallo del prototipo que impide seguir debe registrarse como incidente técnico, no como incapacidad del participante.

No penalicen una ruta alternativa si conduce al resultado correcto. Si el equipo esperaba tres clics y la persona usa cinco, la tarea puede ser eficaz aunque menos eficiente. El criterio de éxito se fija antes de las sesiones para que no cambie según el resultado que se obtiene.

> La consigna se lee sin explicar la interfaz. El criterio interno de finalización permite evaluar rutas distintas y documentar la incertidumbre.

Referencias: [4], [8] (ver bibliografía).

13 / MODERACIÓN

Think Aloud: escuchar mientras la persona actúa

La técnica de pensamiento en voz alta permite conocer expectativas y dudas en el momento en que se manifiestan.

En un protocolo concurrente de *Think Aloud*, la persona realiza la tarea y verbaliza lo que está pensando: qué busca, qué reconoce, qué espera de un control o qué le genera incertidumbre. No se le solicita que justifique cada acción ni que explique cómo debería diseñarse el producto. Se intenta acceder a parte de su proceso mientras ocurre, sin transformar la tarea en una entrevista guiada.

La posibilidad de escuchar el razonamiento ayuda a interpretar una pausa o un clic equivocado. Sin embargo, una verbalización no es una descripción completa de los procesos mentales. Hay acciones automáticas difíciles de explicar y personas que verbalizan más que otras. Además, hablar puede aumentar el tiempo empleado y modificar la forma de resolver una tarea; por eso esta modalidad es particularmente útil para evaluación cualitativa, no para medir tiempos naturales de uso con precisión.

Cómo introducir la técnica

Guion de inicio propuesto

«Estamos estudiando la interfaz, no tus conocimientos. Usala como lo harías normalmente. Mientras lo hacés, contanos lo que estás buscando, lo que esperás que pase y cualquier duda que aparezca. No hace falta que nos expliques todo ni que intentes ayudarnos a quedar bien».

Puede hacerse un breve ensayo con una actividad diferente (por ejemplo, localizar un icono en una pantalla neutra). Durante la tarea, si la persona deja de hablar, se espera unos segundos. Solo cuando el silencio impide comprender lo que ocurre se utiliza un recordatorio discreto: «¿Qué estás pensando ahora?». Las preguntas que exigen reflexionar extensamente se reservan para el cierre.

> El pensamiento en voz alta complementa la observación. No es obligatorio que el usuario verbalice continuamente ni se debe presionarlo; puede preferir hacer pausas o necesitar adaptaciones.

Referencias: [4], [24], [25] (ver bibliografía).

14 / MODERACIÓN

Roles, ubicación y neutralidad

Diseñar la situación de prueba es tan importante como diseñar la consigna.

El moderador recibe a la persona, presenta la actividad y administra las intervenciones. Debe mantener una actitud respetuosa y neutral, evitando asentir selectivamente, corregir respuestas o celebrar determinadas acciones. Si el participante se equivoca, el moderador no debe rescatarlo de inmediato: primero registra el problema y observa si consigue recuperarse. Una ayuda necesaria para continuar tiene que quedar anotada.

El observador registra las acciones, verbalizaciones y condiciones de la sesión. En una prueba presencial puede ubicarse oblicuamente, fuera de la línea de mirada principal, con visión suficiente de las manos y la pantalla, sin invadir el espacio personal. Su presencia debe estar informada. Las expresiones corporales pueden aportar contexto, pero no prueban por sí mismas un estado mental. Se registra «permanece 12 segundos sin tocar la pantalla», no «está frustrado».



Esquema de ubicación sugerida. La comodidad, el consentimiento y el acceso visual determinan la disposición final.

Si se trabaja individualmente

Una misma persona puede moderar y tomar notas mediante una plantilla corta. Es mejor registrar unos pocos hechos precisos que intentar transcribir todo y perder el seguimiento de la tarea. La grabación, cuando se autoriza, puede facilitar la revisión posterior, pero no es requisito del práctico.

> El moderador está allí para obtener evidencia, no para demostrar lo bien que funciona su propio diseño.

Referencias: [4], [25], [26] (ver bibliografía).

15 / MODERACIÓN

Cuándo intervenir y cómo formular preguntas

El objetivo de una intervención es comprender o mantener la sesión, no orientar a la solución.

Situación	Intervención neutral	Evitar
Silencio prolongado	¿Qué estás pensando en este momento?	¿Estás buscando el saldo?
Duda antes de tocar	¿Qué esperás que ocurra?	Ese botón te lleva al detalle.
Pregunta: ¿Está bien así?	¿Qué harías si estuvieras usando esto por tu cuenta?	Sí, vas bien; seguí por ahí.
Bloqueo persistente	Registrar el bloqueo y, según protocolo, ofrecer una ayuda explícita.	Hacer clic por la persona sin anotarlo.
Error técnico	Detener el tiempo, describir el incidente y reiniciar si corresponde.	Contabilizarlo automáticamente como error de usuario.

Ayuda y criterio de finalización

Antes de empezar definan cuándo finalizar un intento: por ejemplo, si se alcanza el límite acordado, la persona quiere detenerse o existe un bloqueo técnico. Si el objetivo es observar el siguiente paso, puede darse una pista después de registrar la falla inicial. Desde ese momento, el resultado ya no constituye un éxito sin ayuda. Es posible aprender del resto del recorrido sin falsear la métrica.

Después de completar o abandonar

En el cierre se exploran las expectativas y las causas posibles de dificultades mediante preguntas abiertas. Conviene solicitar primero una valoración individual de facilidad, si se utiliza SEQ, y luego conversar para no condicionarla. Preguntas como «¿qué esperabas encontrar?» son mejores que «¿te pareció confuso el menú?», porque no presuponen una valoración. Un comentario puede sugerir una interpretación que debe contrastarse con el registro.

Diálogo ficticio

P: «No sé si ya lo pedí». **M:** «¿Qué te hace pensar eso?». **P:** «Sigo viendo el mismo botón y el saldo no cambió». **Registro:** duda en confirmación; la interfaz permanece en el detalle. El moderador no indica qué falta ni explica el comportamiento.

Referencias: [4], [24], [25] (ver bibliografía).

16 / ÉTICA Y MODALIDADES

Consentimiento, privacidad y contexto de prueba

Una evaluación debe ser voluntaria, respetar a las personas y minimizar los datos recogidos.

El consentimiento informado se obtiene antes de observar o registrar la interacción. Hay que comunicar, con lenguaje accesible, el propósito académico, la duración, las actividades, quién accederá a los resultados, si habrá grabación y cómo se conservarán los datos. La persona puede negarse a participar o dejar la sesión sin consecuencias. La autorización para participar no implica automáticamente permiso para grabar audio, video o pantalla.

Utilicen cuentas y datos ficticios, especialmente en tareas transaccionales. No se necesita una contraseña real para evaluar una pantalla de confirmación. Identifiquen participantes con códigos (P01, P02) y separen cualquier registro de consentimiento de la planilla de resultados. No incluyan correos, fotografías ni datos sensibles en un informe académico si no son necesarios. Acuerden el resguardo, los plazos de conservación y cómo retirar datos cuando sea posible.

Modelo orientativo de presentación

«La actividad dura aproximadamente 10 a 15 minutos y forma parte de un trabajo académico. Observaremos cómo usás un prototipo con datos ficticios y registraremos notas sin tu nombre. Podés detenerte cuando quieras. No vamos a grabar a menos que lo autorices por separado. ¿Estás de acuerdo en participar?»

Participación accesible

Si la población incluye personas con discapacidad, contemplen sus tecnologías de asistencia habituales, tiempos y formas de comunicación. No confundan una barrera del entorno de prueba con una limitación del usuario. El pensamiento en voz alta puede adaptarse, por ejemplo, para evitar competir con la voz de un lector de pantalla. Registren la adaptación y consideren su efecto sobre la comparación.

> En el TP no es obligatorio grabar. Las notas anónimas y precisas resultan suficientes si documentan la evidencia necesaria.

Referencias: [4], [9], [10], [26], [27] (ver bibliografía).

17 / ÉTICA Y MODALIDADES

Pruebas moderadas, no moderadas y remotas

La elección de modalidad debe obedecer al objetivo del estudio y no a la disponibilidad de una herramienta.

Una prueba **moderada** cuenta con una persona que acompaña la sesión, controla el protocolo y puede aclarar la consigna sin revelar la solución. Es valiosa cuando todavía no se sabe por qué ocurren las dificultades. En una prueba **no moderada**, el participante sigue las instrucciones por su cuenta; los registros se recogen mediante un sistema o un procedimiento previamente preparado. Requiere que la ruta funcione y la tarea tenga un resultado claro, porque no se puede repreguntar en el momento.

Modalidad	Qué aporta	Qué controlar
Presencial moderada	Observación directa, contexto y posibilidad de repreguntar.	Influencia del equipo, espacio y dispositivo.
Remota moderada	Acceso a perfiles dispersos y uso del equipo habitual.	Conexión, permisos, audio y pantalla compartida.
No moderada remota	Más intentos con instrucciones estandarizadas.	Calidad del reclutamiento, funcionamiento y registros incompletos.
No moderada presencial	Autonomía en un entorno preparado.	Instrucciones y captura fiable de resultados.

En remoto, hagan una comprobación técnica inicial, distingan el tiempo de setup del tiempo de tarea y acuerden qué pantalla se compartirá. Eviten solicitar control remoto del dispositivo si no es necesario. Una persona puede dudar porque no sabe manejar la herramienta de videollamada; esa dificultad no debe atribuirse sin análisis al prototipo que se está evaluando.

> Para el TP 5 se recomienda una sesión moderada, presencial o remota, de 10–15 minutos por participante. Permite aplicar Think Aloud sin depender de plataformas especializadas.

Referencias: [4], [10], [11] (ver bibliografía).

18 / GUÍA DE CAMPO

Una sesión de 10–15 minutos, de principio a fin

El guion mantiene condiciones comparables sin convertir la prueba en un interrogatorio rígido.

Momento	Acción del moderador	Registro
0–2 min · Inicio	Presentar objetivo, consentimiento y carácter voluntario.	ID, autorización, dispositivo.
2–3 min · Contexto	Preguntar si realizó alguna operación semejante.	Experiencia relevante, sin enseñar el diseño.
3–4 min · Consigna	Explicar Think Aloud; leer la tarea y entregar datos.	Versión y criterio observacional.
4–11 min · Tarea	Observar en silencio; recordar que verbalice si hace falta.	Tiempo, acciones, dudas, errores y ayudas.
11–13 min · Cierre	Preguntar facilidad y expectativas; agradecer.	SEQ opcional, comentarios finales.

Los tiempos son una propuesta para una operación acotada y no una duración obligatoria de toda prueba profesional. Si la persona termina antes, no se agregan tareas improvisadas solo para ocupar el tiempo. Si no completa, se documenta la causa y se da cierre de acuerdo con el límite definido. El orden de actividades y las condiciones iniciales deberían mantenerse estables para todos los participantes de una ronda.

Antes de recibir a la siguiente persona

Restablezcan el saldo ficticio, el estado inicial y el historial; anoten cualquier modificación en el prototipo. El observador puede completar las notas mientras la sesión está fresca. No cambien etiquetas tras cada participante sin identificar una nueva versión: hacerlo impide interpretar a quién se expuso a cada diseño. Si se necesita corregir un bloqueo, documenten el cambio y separen las observaciones por versión.

> Una prueba breve, bien preparada y bien registrada, es más útil que una sesión extensa en la que se explica la interfaz o se modifica la consigna continuamente.

Referencias: [4], [23], [25], [26] (ver bibliografía).

19 / EVIDENCIA

Observar, registrar e interpretar

El material principal de una investigación de usabilidad son las acciones observadas en contexto.

El registro debe permitir que otra persona reconstruya lo que ocurrió. Una nota como «no entendió el sistema» es una interpretación difícil de verificar. En cambio, consignar que abrió una sección, volvió atrás, permaneció varios segundos sin actuar y formuló una pregunta proporciona una secuencia concreta. Al redactar citas textuales, distingan lo literal de una paráfrasis; si no pudieron anotar las palabras exactas, identifíquenlas como resumen.

Tiempo	Hecho observable (ejemplo ficticio)	Interpretación tentativa
00:20	P02 toca «Mi cuenta».	Espera consultar allí sus puntos.
00:34	Vuelve al inicio y dice que busca beneficios.	La ruta de acceso no es evidente.
01:17	P02 pulsa «Canjear». Nada cambia.	No se comunica el estado.
01:22	Pregunta «¿ya está?».	Incertidumbre sobre resultado.
01:24	Pulsa nuevamente el botón.	Posible riesgo de duplicación.

Qué registrar además de los clics

Es útil anotar ruta seguida, cantidad de intentos, pausas significativas, mensajes visibles, ayudas del moderador y circunstancias técnicas. Las verbalizaciones complementan esa historia: qué esperaba que sucediera y qué creyó que sucedió. Un gesto o una expresión facial se describe como conducta visible, sin adjudicarle emociones no expresadas. El contexto del dispositivo puede explicar problemas que no se reproducen en otro entorno.

Después de la sesión, revisen cada nota y separen incidente, evidencia, posible interpretación y nivel de certeza. Los hechos deben conservarse aunque el equipo decida descartar una interpretación. Este método facilita volver a los registros cuando aparecen nuevas observaciones o alguien cuestiona una conclusión.

> La evidencia pertenece al momento y a la versión del prototipo que se observó. Un problema deducido por el equipo no debe presentarse como una cita del usuario.

Referencias: [4], [26] (ver bibliografía).

20 / MUESTRA

Cinco usuarios: alcance y límites de la evidencia

El número de participantes se define según el objetivo: descubrir problemas no es lo mismo que estimar una proporción poblacional.

En estudios formativos, las primeras sesiones suelen revelar dificultades frecuentes; agregar participantes similares puede producir rendimientos decrecientes. Nielsen y Landauer fundamentaron esta idea con un modelo de descubrimiento: si cada participante tiene una probabilidad p de encontrar un problema y los intentos son independientes, la probabilidad de observarlo al menos una vez con n participantes es:

$$P(\text{detectar}) = 1 - (1 - p)^n$$

Con p = 0,31, valor ilustrativo usado por Nielsen, cinco participantes darían aproximadamente 84 % de probabilidad de detectar un problema de esa detectabilidad; con dos, alrededor de 52 %. No significa que cinco personas detecten el 85 % de los problemas de cualquier aplicación. Las probabilidades reales difieren según problema, perfil, tarea y contexto. Un caso raro puede pasar inadvertido incluso después de muchas sesiones.

Probabilidad por persona	2 participantes	5 participantes
31 %	52,4 %	84,4 %
10 %	19,0 %	41,0 %

Si dos de cinco participantes fallan, se puede decir que dos personas de esa muestra fallaron en esas condiciones. No se puede inferir que el 40 % de toda la población fallará. Para estimar una tasa con precisión, comparar alternativas estadísticamente o estudiar perfiles heterogéneos se necesita otro diseño de muestreo. La calidad del reclutamiento importa tanto como la cantidad.

> **Condición del práctico:** mínimo dos usuarios por grupo, ajenos al diseño y próximos al público objetivo. Dos alcanzan para ejercitar la técnica, no para certificar la usabilidad. Si se pueden sumar más, especialmente hasta cinco o más participantes pertinentes, se recomienda hacerlo.

Referencias: [12], [13], [28], [32] (ver bibliografía).

21 / MÉTRICAS

Eficacia, eficiencia y errores

Medir consiste en asignar un significado operativo a una observación, no solo en contar clics.

Las métricas de tarea deben definirse antes de observar los resultados. El equipo decide qué se considera un intento válido, una finalización correcta, una intervención sustantiva y un error. Esto permite comparar los registros sin cambiar el criterio cuando alguien sigue una ruta inesperada. Los indicadores describen el desempeño bajo el protocolo empleado, no una propiedad universal del producto.

Tasa de éxito

Si se exige completar la tarea sin intervención, el numerador incluye solamente los éxitos independientes. El denominador corresponde a intentos evaluables de la tarea. Las finalizaciones asistidas y los abandonos se muestran aparte; las interrupciones técnicas se registran y se justifica si se excluyen. Nunca se oculta una exclusión porque perjudica el resultado.

Éxito (%) = éxitos independientes / intentos válidos × 100

Tiempo en tarea

Definan un comienzo observable (por ejemplo, cuando la persona inicia la interacción tras leer la consigna) y un final (resultado, abandono o límite). Informen los tiempos por separado para éxitos y fallas; mezclar ambos puede producir un indicador engañoso, ya que abandonar rápido no es ser eficiente. En pruebas Think Aloud los tiempos son orientativos: verbalizar y recibir intervenciones puede alterar la duración.

Errores y oportunidades

Un error es una acción que se aparta del objetivo de manera relevante según la definición acordada. Explorar una ruta alternativa exitosa no es necesariamente un error. Se puede informar errores por intento o, si existen oportunidades bien definidas y comparables, errores divididos por oportunidades. Si el denominador no puede definirse de manera fiable, es preferible presentar el recuento y su descripción.

> La medida debe conservar su contexto: «dos reintentos de confirmación en cinco sesiones» explica más que un porcentaje de errores sin denominador.

Referencias: [14], [29] (ver bibliografía).

22 / MÉTRICAS

Ejemplo resuelto: cinco participantes

Datos ficticios de una ronda moderada. La tarea consiste en obtener una entrada y reconocer que quedó registrada.

El inicio del cronómetro es la primera acción. El final es el reconocimiento del resultado o el cierre del intento. El equipo registra como error elegir una opción no correspondiente o repetir una acción de confirmación; la exploración sin consecuencia no se contabiliza por sí sola. Las ayudas sustantivas impiden clasificar el resultado como éxito independiente.

ID	Resultado	Tiempo	Errores	SEQ	Nota relevante
P01	Éxito	72 s	0	6	Reconoce comprobante.
P02	Éxito	95 s	1	4	Repite confirmación.
P03	No completa	180 s	1	2	Elige sección equivocada; abandona buscando saldo.
P04	Éxito asistido	110 s	1	3	Elige opción incorrecta; recibe pista de navegación.
P05	Éxito	88 s	0	5	Duda si terminó; luego reconoce resultado.

3/5
60 % éxito independiente

88 s
mediana de éxitos

4/7
promedio SEQ

Cálculos e interpretación

Éxito: 3 / 5 × 100 = 60 %. Si se informa el éxito incluyendo asistencia, 4 / 5 = 80 %, pero debe llevar esa etiqueta. **Tiempo:** ordenar los tres éxitos independientes (72, 88, 95); mediana = 88 s, media = 85 s. **Errores:** 3 en total, 3 / 5 = 0,6 por intento; tres participantes tuvieron al menos uno. **SEQ:** (6 + 4 + 2 + 3 + 5) / 5 = 4.

Los resultados señalan dos zonas para investigar: localización del saldo y confirmación. P02 y P05 expresan incertidumbre al finalizar, aunque ambos completaron; P03 no llegó a esa etapa. En v1 el comprobante puede encontrarse en el historial, pero no se muestra automáticamente tras confirmar. El análisis de la confirmación tiene que considerar únicamente a quienes estuvieron expuestos a ella. Los datos permiten orientar una mejora, no establecer el desempeño de toda la población.

> Ejemplo de conclusión válida: en esta ronda, dos participantes manifestaron dudas sobre el resultado; proponemos mejorar la confirmación y repetir la tarea.

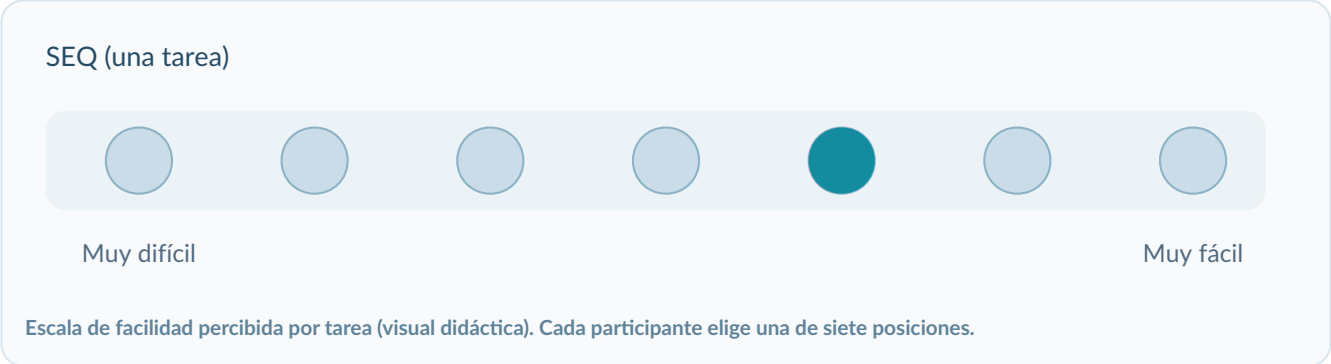
Referencias: [14], [15], [29] (ver bibliografía).

23 / PERCEPCIÓN

SEQ: facilidad percibida por tarea

La Single Ease Question registra la percepción inmediata de dificultad, después de intentar una tarea.

La observación permite medir si una operación se completó y con qué esfuerzo; el cuestionario registra cómo fue vivida. La pregunta SEQ se aplica inmediatamente después de cada intento, incluso si la persona no lo completó. El participante selecciona una posición en una escala de siete puntos, desde muy difícil hasta muy fácil. La pregunta se formula de forma neutral y se mantiene igual durante el estudio.



Escala de facilidad percibida por tarea (visual didáctica). Cada participante elige una de siete posiciones.

Pregunta orientativa en español

«En general, ¿qué tan difícil o fácil te resultó realizar esta tarea?»

1 = muy difícil · 4 = punto medio · 7 = muy fácil. Si se utiliza una versión estandarizada en un estudio formal, deben mantenerse la redacción validada, el orden de la escala y las condiciones de administración.

Una tarea completada puede recibir un puntaje bajo porque exigió demasiada concentración. Una tarea fallida puede recibir un puntaje alto si la persona creyó erróneamente que había alcanzado el resultado. Esa discrepancia no invalida la percepción: aporta una pista sobre el modelo mental y sobre la comunicación del estado del sistema. Por eso el SEQ se analiza junto a éxito, tiempo y evidencia cualitativa.

En el ejemplo de cinco participantes, la mediana es 4, el rango va de 2 a 6 y la media es 4. Con dos participantes puede ser más informativo mostrar las dos respuestas que comunicar solo un promedio. En ningún caso el puntaje es una medida directa de carga cognitiva ni una nota de aprobación universal.

> El SEQ es opcional en el TP 5. Su uso permite practicar medición de percepción sin extender excesivamente la sesión.

Referencias: [15], [29] (ver bibliografía).

24 / PERCEPCIÓN

SUS: evaluación global de la usabilidad percibida

La System Usability Scale sintetiza la percepción de uso de un sistema mediante diez ítems.

El SUS fue desarrollado por John Brooke y publicado en 1996. Se administra después de que la persona haya tenido una experiencia suficiente con el sistema. A diferencia del SEQ, que pregunta por una tarea específica, SUS recoge una impresión global sobre facilidad, complejidad, consistencia, aprendizaje y confianza. El instrumento utiliza diez afirmaciones y cinco opciones de respuesta por ítem (de desacuerdo a acuerdo).

En la versión clásica los ítems impares están formulados positivamente y los pares negativamente. Para obtener el puntaje se transforma cada respuesta: a los impares se les resta 1; a los pares se les resta su respuesta a 5. Cada contribución queda entre 0 y 4. La suma se multiplica por 2,5 para expresarla en una escala de 0 a 100.

$$SUS = 2{,}5 \times [\Sigma(\text{impares} - 1) + \Sigma(5 - \text{pares})]$$

Ítem	1	2	3	4	5	6	7	8	9	10
Respuesta	4	2	5	2	4	2	4	1	4	2
Aporte	3	3	4	3	3	3	3	4	3	3

Ejemplo ficticio: la suma de aportes es 32; 32 × 2,5 = **80**. El número no significa 80 % de usabilidad ni 80 % de tareas completadas. Puede compararse con datos de referencia cuando las condiciones y la versión son compatibles, pero el punto de corte no es un criterio absoluto de aprobación de un producto.

Uso responsable del instrumento

La puntuación requiere los diez ítems completos y su codificación correcta. No se debe improvisar una traducción para un estudio formal ni eliminar las preguntas que resulten incómodas. Consulten la escala original y una traducción validada para su población. En el TP 5 no se exige SUS: con una tarea y dos usuarios se priorizan la observación y las métricas básicas.

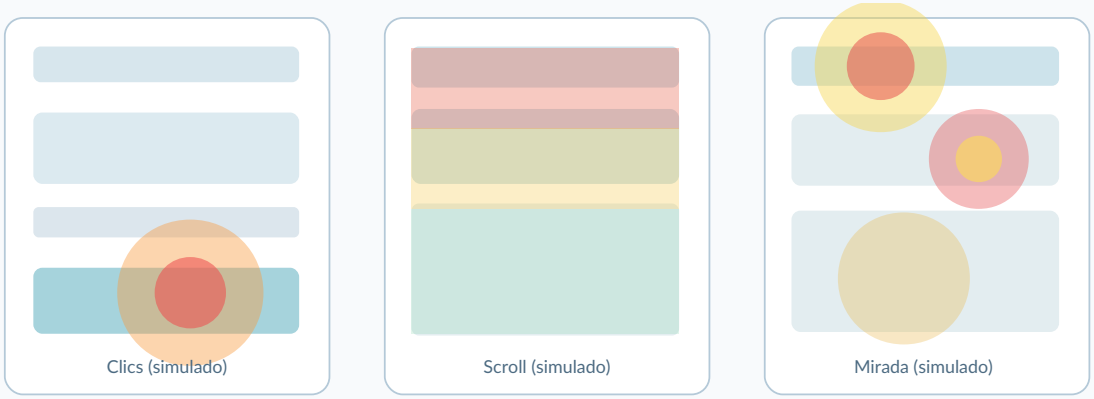
> SUS y SEQ son indicadores de percepción, cuantificados mediante respuestas; los comentarios abiertos y las observaciones cualitativas permiten investigar por qué obtuvieron esos valores.

Referencias: [16], [30], [33] (ver bibliografía).

25 / MAPAS DE CALOR

Qué representan y qué no representan

Un mapa de calor es una representación espacial de una señal agregada. Su interpretación depende de cómo se obtuvo esa señal.



Tres visualizaciones sintéticas y conceptuales; no son resultados de un estudio real.

Tipo	Datos que utiliza	Cuidado interpretativo
Clics o toques	Coordenadas o elementos interactuados.	Muchos clics pueden significar interés o confusión.
Scroll	Profundidad o proporción de sesiones que llegan a una zona.	Llegar no equivale a leer ni comprender.
Eye tracking	Posición de mirada y fijaciones capturadas por sensores.	Mirar no demuestra entender, preferir o recordar.
Predicción de IA	Estimaciones a partir de capturas u otras entradas.	No registra miradas ni acciones reales.

Los mapas de clics y scroll suelen construirse con eventos de productos implementados. Permiten estudiar patrones bajo determinadas resoluciones, dispositivos y rutas. La agregación puede ocultar diferencias importantes: una misma coordenada puede corresponder a elementos distintos si cambia el diseño responsive. Antes de comparar mapas, revisen período, cantidad de sesiones, dispositivo, estado de pantalla y versión.

El seguimiento ocular requiere calibración y un procedimiento adecuado. Una zona con fijaciones intensas puede atraer atención porque es relevante o porque resulta difícil de interpretar. Para saber cuál es la causa se necesita relacionar el mapa con la tarea y otras observaciones. El color rojo no es un indicador universal de buena o mala experiencia.

> Los tres mapas de la ilustración son esquemas simulados; sirven para explicar la técnica, no constituyen evidencia de usuarios.

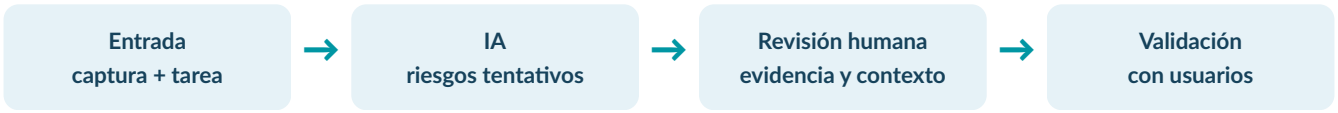
Referencias: [17], [18], [31] (ver bibliografía).

26 / IA APLICADA

Evaluación predictiva y asistencia de inteligencia artificial

Una herramienta predictiva puede señalar riesgos visuales o sugerir hipótesis, pero no observar por sí sola a los destinatarios del producto.

Las herramientas de predicción de atención pueden analizar una captura para estimar qué zonas resultarán visualmente salientes. Otras utilizan modelos de lenguaje y visión para identificar posibles problemas de consistencia, etiquetado o jerarquía. El resultado depende del modelo, la información de entrada y el contexto que el sistema puede representar. No equivale a un registro de una sesión con personas ni verifica la función real del software.



Aplicación antes y después de una prueba

Antes de evaluar, la IA puede revisar el guion y detectar formulaciones inductivas, proponer casos límite o ayudar a redactar una lista de observación. Después, puede organizar notas anónimas y sugerir agrupamientos de hallazgos. En ambos casos, el investigador debe revisar los resultados contra la información original. No se deben inventar citas, convertir hipótesis en hechos ni completar datos ausentes con respuestas plausibles.

Prompt para una revisión asistida

«Analizá esta captura del flujo de canje y el objetivo de la tarea. Identificá posibles obstáculos para un usuario nuevo. Para cada uno indicá qué se puede verificar en la captura, cuál es tu supuesto y qué observación permitiría contrastarlo. No atribuyas acciones o emociones a personas que no fueron observadas».

Los llamados usuarios sintéticos pueden servir para explorar escenarios o practicar moderación, pero no reemplazan el reclutamiento. Asimismo, antes de compartir notas o grabaciones con servicios externos, revisen consentimiento, privacidad y condiciones de tratamiento de datos. El TP exige pruebas con personas reales; la IA solo acompaña el análisis.

> Distinguir en el informe: observación real, inspección humana, predicción de IA y ejemplo ficticio. Son fuentes diferentes y no tienen el mismo valor probatorio.

Referencias: [18], [19], [20] (ver bibliografía).

27 / SÍNTESIS

Del registro individual a un hallazgo

La síntesis transforma episodios aislados en problemas documentados, sin borrar sus diferencias.

Al terminar la ronda, revisen las notas y asignen identificadores a incidentes relevantes. Agrupen los que se refieren a una misma fricción y separen aquellos que tienen consecuencias o causas distintas. No agrupen solo porque dos personas hicieron clic en el mismo botón: una puede haber buscado información y otra haber intentado repetir una operación. Conservar el objetivo de cada participante facilita interpretar el comportamiento.

Caso ficticio	Evidencia	Hallazgo
P02 · 01:22	Pregunta «¿ya está?» y repite el botón.	H-01: resultado insuficientemente claro.
P05 · final	Expresa duda, luego encuentra comprobante.	H-01: la confirmación no se reconoce de inmediato.
P03 · 03:00	Abandona al buscar sus puntos.	H-02: dificultad para localizar saldo.
P04 · durante	Necesita pista para llegar al catálogo.	H-03: ruta de acceso poco evidente.

Redactar un problema sin imponer una solución

Eviten escribir «falta un botón verde» como problema. Esa frase ya propone una alternativa. Es mejor explicar que, tras confirmar, la persona no logra distinguir si la operación se registró, lo que genera dudas o reintentos. Así pueden considerarse mensajes, un cambio de estado, un comprobante o una actualización del saldo. La decisión de diseño debe evaluarse después.

Un hallazgo puede tener respaldo en una única sesión si el impacto es significativo. Indiquen cuántas personas lo experimentaron y cuántas estuvieron expuestas al paso, sin convertir una muestra reducida en una estimación general. Si una causa no está establecida, preséntenla como hipótesis. Los resultados negativos, contradicciones y rutas exitosas también son evidencia.

> Formato útil: contexto + conducta observable + consecuencia para la tarea + evidencia identificable + posible explicación.

Referencias: [4], [5], [26] (ver bibliografía).

28 / PRIORIZACIÓN

Decidir qué corregir primero

La priorización convierte los hallazgos en decisiones de trabajo. No consiste en ordenar opiniones por cantidad de votos.

Se parte de la severidad, considerando el impacto y la frecuencia observada o estimada. A diferencia de la severidad, la prioridad también incorpora el riesgo de negocio, las dependencias y el costo de una modificación. Un problema poco frecuente puede merecer atención inmediata si produce una consecuencia irreversible. Un detalle visual repetido puede resolverse después si no afecta la tarea principal.

ID	Impacto y evidencia en el caso	Severidad	Decisión
H-01	P02 repite confirmación; P05 duda del resultado. Riesgo de duplicación no comprobado.	3 provisional	Alta: revisar estado y confirmación.
H-02	P03 no completa por no ubicar saldo.	3	Alta: mostrar saldo en contexto.
H-03	P04 requiere ayuda para llegar al catálogo.	3	Revisar etiqueta y jerarquía de acceso.
H-04	Inspección: alineación desigual, sin efecto en tarea demostrado.	1	Baja: ajustar con el sistema visual.

Cuando hay dos problemas con severidad similar, expliciten por qué se prioriza uno. En este caso, la confirmación está al final de la operación y puede provocar reintentos; una falla de saldo impide acceder al beneficio. Ambos son relevantes. Si los recursos son limitados, el equipo puede seleccionar uno primero y dejar el otro documentado, con una justificación. No es correcto afirmar que los demás quedaron resueltos porque no se trabajaron en esta iteración.

Ficha de decisión

H-01 · Prototipo v1. Probar un estado de procesamiento, deshabilitación temporal y confirmación persistente con comprobante. **Motivo:** los participantes no distinguen si se registró el canje. **Verificación:** observar reconocimiento espontáneo y ausencia de reintentos en una nueva ronda.

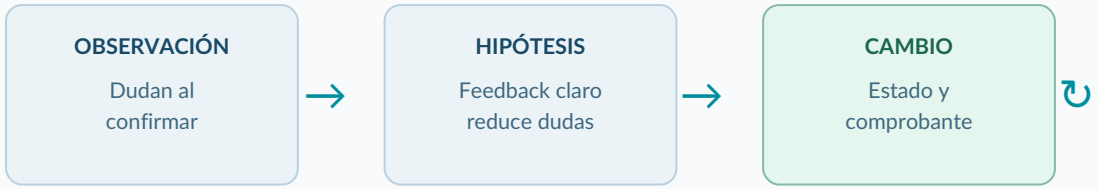
> Se prioriza el problema, no la ocurrencia de una solución estética. Las modificaciones deben poder vincularse con evidencia.

Referencias: [5], [21] (ver bibliografía).

29 / ITERACIÓN

Validar hipótesis mediante evidencia

Iterar significa modificar una versión y volver a contrastar las decisiones con el objetivo de la tarea.



La mejora debe volver a observarse. No basta con asumir que el cambio funciona.

La hipótesis debe declarar el cambio, el comportamiento esperado y el modo de verificarlo. En el caso de H-01: si el sistema presenta un estado de procesamiento y una confirmación persistente, esperamos que las personas reconozcan que el canje fue realizado sin repetir la acción. El resultado puede observarse mediante éxito independiente, verbalizaciones y número de reintentos; no basta con preguntar si la pantalla nueva es más linda.

Cómo organizar un retest

Conserven la tarea, el criterio de éxito, los datos ficticios y condiciones comparables. Registren la versión evaluada. Usar nuevos participantes de perfil similar reduce la influencia del aprendizaje de la primera versión. Si se utilizan las mismas personas, aclaren que ya conocen el recorrido y pueden recordar dónde actuar. En muestras pequeñas una comparación descriptiva aporta indicios, no una demostración estadística de superioridad.

Estado	Ejemplo de registro
v1 · evidencia	P02 repite confirmación; P05 duda.
v2 · cambio	Estado pendiente, confirmación visible y comprobante.
Retest · pendiente	Observar si se reconoce el resultado sin ayuda y sin reintentos.

Si no se puede realizar otra ronda durante el cursado, documenten la modificación y el criterio que utilizarían para comprobarla. El TP exige aplicar al menos una mejora; no obliga a declarar una mejora validada sin datos. La Unidad 6 retomará esta decisión en forma de especificación y criterios de aceptación.

> Diferenciar «cambio implementado» de «cambio evaluado» es fundamental para mantener la trazabilidad y la honestidad del informe.

Referencias: [1], [4], [21] (ver bibliografía).

30 / INTEGRACIÓN

El caso completo y su relación con el TP 5

Una prueba didáctica puede recorrer el ciclo de evaluación entero sin exigir una implementación real.

El prototipo de canje comienza en la pantalla de catálogo, pasa por el detalle de una entrada y termina en confirmación. La investigación de la Unidad 1 sugirió que los usuarios quieren saber qué beneficios pueden utilizar con sus puntos. La evaluación heurística identifica riesgos de feedback y de visibilidad del saldo; el recorrido cognitivo anticipa dificultades en las etiquetas. Esos resultados orientan la tarea, no reemplazan las sesiones.

Etapa	Aplicación
Preparación	Prototipo v1, saldo 1.800, entrada 1.200, tarea y éxito definidos.
Sesiones	Participantes reales y pertinentes. Think Aloud moderado con consentimiento.
Registro	Resultado, tiempo, errores, ayudas, dudas y citas anónimas.
Síntesis	Relacionar evidencia con H-01, H-02 y H-03; revisar la gravedad.
Iteración	Aplicar al menos una mejora y conservar el enlace entre problema y cambio.
Entrega	Informe breve y prototipo actualizado, con evidencia y límites.

Alcance oficial del práctico

Se exige un mínimo de dos participantes por grupo, preferentemente más si es posible. La tarea debe recorrer el diseño de la Unidad 4 y las sesiones duran aproximadamente 10–15 minutos. El informe de 2–3 carillas resume metodología, resultados, problemas respaldados por evidencia y mejoras. El prototipo se entrega actualizado en Figma o HTML; no se exige backend. El enunciado oficial prevalece sobre cualquier sugerencia metodológica ampliada de este apunte.

En el ejemplo de las cinco sesiones no hay usuarios reales: los datos se construyeron exclusivamente para mostrar el método. En el trabajo académico, las observaciones deben corresponder a personas efectivamente convocadas. Si faltan datos o una tarea no puede ejecutarse, documenten la limitación en vez de completarlos con estimaciones ficticias.

> El objetivo didáctico es aprender a producir evidencia y tomar decisiones justificadas, incluso cuando la interfaz no funciona como se esperaba.

31 / PLANTILLAS

Instrumento de planificación y registro

Modelos reutilizables para el TP: pueden adaptarse al proyecto sin perder los campos necesarios para interpretar las sesiones.

Plan previo de la prueba

Campo	Completar antes de convocar
Proyecto / versión	_____
Pregunta / hipótesis	_____
Perfil y dispositivo	_____
Escenario y datos de prueba	_____
Criterio de éxito / ayuda / cierre	_____
Definición de error y cronometraje	_____

Ficha anónima individual

ID: _____ · Versión: _____ · Modalidad: _____ · Consentimiento: sí / no.
Resultado: independiente / asistido / no completa / incidente técnico.
Tiempo: _____ · Errores: _____ · SEQ (opcional): _____.

Momento	Conducta o cita literal	Interpretación tentativa

Consolidación de hallazgos

ID	Evidencia / usuarios	Impacto	Severidad	Cambio y verificación
H-01				
H-02				
H-03				

> Al completar la ficha, no incluyan datos personales innecesarios. Si un intento se interrumpe por un incidente técnico, dejen registro del motivo y expliquen su tratamiento en el análisis.

32 / ESTUDIO

Preguntas para revisar lo aprendido

Resuelvan estas situaciones antes de utilizar el prototipo con participantes. Las respuestas orientativas permiten comprobar sus argumentos.

1. Fidelidad e inspección

Un prototipo tiene diseño visual detallado, pero el botón de confirmar no navega. ¿Es apropiado para evaluar éxito de tarea?

Respuesta orientativa: no, si ese control es necesario para finalizar. Puede servir para estudiar jerarquía y comprensión de la etiqueta. Hay que construir o simular el resultado antes de medir éxito.

2. Moderación

La persona pregunta dónde están sus puntos. El moderador le indica el menú. ¿Cómo se registra?

Respuesta orientativa: registrar la duda, la pista y el resultado. Si después completa, es una finalización asistida, no un éxito independiente. La intervención permite aprender del resto de la tarea.

3. Tamaño de muestra

Uno de dos participantes no encuentra el saldo. ¿Qué podemos afirmar?

Respuesta orientativa: una persona de la muestra encontró una dificultad en esas condiciones; merece investigación según su impacto. No se puede afirmar que la mitad de todos los usuarios fallará.

4. Métricas y percepción

Una tarea termina correctamente, pero obtiene SEQ 2/7. ¿Hay una contradicción?

Respuesta orientativa: no necesariamente. La eficacia y la percepción de esfuerzo son dimensiones diferentes; interesa revisar errores, tiempos y verbalizaciones para interpretar el resultado.

5. Iteración

Una IA predice que el botón se verá mejor después del cambio. ¿Puede declararse validado?

Respuesta orientativa: no. La predicción ayuda a formular una hipótesis; se necesita observar el comportamiento con el nuevo diseño. Si no se retesteó, la validación queda pendiente.

> Para responder, fundamenten con los conceptos de la unidad, no únicamente con una definición memorizada.

BIBLIOGRAFÍA / FUENTES DE REFERENCIA

Referencias y lecturas complementarias

Métodos de validación, pruebas de usabilidad y aspectos de moderación.

Se priorizaron fuentes metodológicas reconocidas, recursos profesionales y referencias originales de las técnicas. Las herramientas y criterios deben leerse según la pregunta de investigación, el tipo de producto y el contexto de uso.

[1] Nielsen Norman Group (2022, revisión 2026). When to Use Which UX Research Methods.
https://www.nngroup.com/articles/which-ux-research-methods/

[2] Nielsen Norman Group. UX Prototypes: Low Fidelity vs. High Fidelity.
https://www.nngroup.com/articles/ux-prototype-hi-lo-fidelity/

[3] Nielsen Norman Group (2024, revisión). 10 Usability Heuristics for UI Design.
https://www.nngroup.com/articles/ten-usability-heuristics/

[4] GOV.UK Service Manual. Using Moderated Usability Testing.
https://www.gov.uk/service-manual/user-research/using-moderated-usability-testing

[5] Jakob Nielsen (1994). Severity Ratings for Usability Problems.
https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

[6] Nielsen Norman Group. Evaluate Interface Learnability with Cognitive Walkthroughs.
https://www.nngroup.com/articles/cognitive-walkthroughs/

[7] Usability Body of Knowledge. Cognitive Walkthrough.
https://usabilitybok.org/cognitive-walkthrough/

[8] Nielsen Norman Group. Task Scenarios for Usability Testing.
https://www.nngroup.com/articles/task-scenarios-usability-testing/

[9] Department for Education, UK. How to Gather and Store Informed Consent.
https://user-research.education.gov.uk/guidance/informed-consent-26/2

[10] GOV.UK. Doing User Research Remotely by Phone or Video Call.
https://www.gov.uk/service-manual/user-research/remote-user-research-phone-video-call

[11] Nielsen Norman Group. Unmoderated User Tests: How and Why to Do Them.
https://www.nngroup.com/articles/unmoderated-usability-testing/

BIBLIOGRAFÍA / FUENTES DE REFERENCIA

Referencias y lecturas complementarias

Muestras, medición, mapas de calor, IA e iteración trazable.

[12] Nielsen Norman Group. Why You Only Need to Test with 5 Users.
https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

[13] Nielsen Norman Group. How Many Test Users in a Usability Study?
https://www.nngroup.com/articles/how-many-test-users/

[14] Usability Body of Knowledge. Usability Testing.
https://www.usabilitybok.org/usability-testing/

[15] MeasuringU. The Evolution of the Single Ease Question (SEQ).
https://measuringu.com/evolution-of-seq/

[16] MeasuringU. Measuring Usability with the System Usability Scale (SUS).
https://measuringu.com/sus/

[17] Nielsen Norman Group. Setup of an Eyetracking Study.
https://www.nngroup.com/articles/eyetracking-setup/

[18] Advances in Human-Computer Interaction (2025). AI in Automated and Remote UX Evaluation.
https://onlinelibrary.wiley.com/doi/full/10.1155/ahci/7442179

[19] Nielsen Norman Group. Synthetic Users: If, When, and How to Use AI-Generated Research.
https://www.nngroup.com/articles/synthetic-users/

[20] Nielsen Norman Group. Accelerating Research with AI.
https://www.nngroup.com/articles/research-with-ai/

[21] Nielsen Norman Group. How to Conduct a Cognitive Walkthrough Workshop.
https://www.nngroup.com/articles/cognitive-walkthrough-workshop/

BIBLIOGRAFÍA / EDICIÓN AMPLIADA

Referencias complementarias

Fundamentos de usabilidad, moderación, métricas y límites de inferencia.

[22] ISO 9241-11:2018. Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts.
https://www.iso.org/standard/63500.html

[23] GOV.UK Service Manual. Plan a round of user research.
https://www.gov.uk/service-manual/user-research/plan-round-of-user-research

[24] Nielsen, J. (2012). Thinking Aloud: The #1 Usability Tool. NN/g.
https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/

[25] Nielsen Norman Group. Checklist for Moderating a Usability Test.
https://www.nngroup.com/articles/usability-checklist/

[26] GOV.UK Service Manual. Taking notes and recording user research sessions.
https://www.gov.uk/service-manual/user-research/taking-notes-and-recording-user-research-sessions

[27] GOV.UK Service Manual. Getting informed consent for user research.
https://www.gov.uk/service-manual/user-research/getting-users-consent-for-research

[28] Nielsen Norman Group. Why 5 Participants Are Okay in a Qualitative Study, but Not in a Quantitative One.
https://www.nngroup.com/articles/5-test-users-qual-quant/

[29] MeasuringU. There are 70+ UX Metrics, Start with These 4.
https://measuringu.com/get-comfortable-with-four-ux-metrics/

[30] MeasuringU. Interpreting Single Items from the SUS.
https://measuringu.com/sus-items/

[31] Nielsen Norman Group. How to Conduct Eyetracking Studies.
https://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/

[32] Nielsen, J. y Landauer, T. K. (1993). A mathematical model of the finding of usability problems. CHI 93, 206–213.
https://doi.org/10.1145/169059.169166

[33] Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. En Usability Evaluation in Industry.
https://doi.org/10.1201/9781498710411-35

Los ejemplos, guiones y registros del apunte son construcciones didácticas propias, salvo donde se indique otra procedencia.