





UNIVERSIDAD DE SEVILLA



Escuela Técnica Superior de Ingeniería Informática  
Departamento de Arquitectura  
y Tecnología de Computadores

**Evaluación Heurística  
de la usabilidad  
en contextos específicos**

TESIS DOCTORAL

Autora:  
Rosa Yáñez Gómez

Directores:  
José Luis Sevillano Ramos  
Daniel Cascado Caballero

Sevilla, abril de 2019



*A Javi, a Maricruz,  
y a los que saben en qué se parecen  
un cuervo y un escritorio*



# Agradecimientos

*«Ítaca te ha dado un viaje hermoso.  
Sin ella no te habrías puesto en marcha.  
Pero no tiene ya más que ofrecerte.*

*Aunque la encuentres pobre, Ítaca de ti no se ha burlado.  
Convertido en tan sabio, y con tanta experiencia,  
ya habrás comprendido el significado de las Ítacas.»  
Cavafis*

La defensa de una Tesis, parafraseando a Derrida<sup>1</sup>, se pospone y excluye por tanto tiempo que cuando finalmente tiene lugar se acaba envolviendo de un carácter de irrealidad e improbabilidad, incluso de improvisación. En este punto de cierre, envuelto en la apariencia de un presente ilusorio, es imprescindible que dedique unas líneas a agradecer su apoyo y paciencia a todos los que me acompañaron en estos años.

En primer lugar, agradecer a mis directores, José Luis Sevillano y Daniel Cascado, su guía constante en este largo, y a veces confuso, camino; su paciencia; su conocimiento.

Agradecer a mi compañero Juan Luis Font su apoyo y trabajo durante esta tesis y, cómo no, su valiosísima amistad. Y a mis compañeras María José Morón y Rocío García el haber podido contar, durante estos años, con su ayuda imprescindible.

A Javi, mi marido, por supuesto, como principal sufridor de los vaivenes de este proyecto, agradecerle su fe inquebrantable, mucho más firme que la mía propia; y tantas cosas que no caben en estas líneas. Y a mi hija Maricruz el haber convocado un terremoto capaz de colocar cada cosa en su lugar y el convertirse en el más ineludible estímulo para perseguir la mejor versión de mí misma; te sigo los pasos deslumbrada, pequeña.

Y a esta Ítaca, que al fin alcanzo, sólo pedirle que me acoja para el descanso durante un tiempo que seguro será breve: apenas el que media hasta embarcar de nuevo buscando nuevas Ítacas.

---

<sup>1</sup> Derrida, J. (1997). *El tiempo de una tesis: deconstrucción e implicaciones conceptuales* (Vol. 3). Anthropos Editorial.





# Resumen

La atención a la usabilidad de los sistemas de información, como requisito no funcional, es especialmente relevante al tratarse de un factor crítico que determina buena parte del volumen de desarrollos correctivos que van a necesitar los productos en su fase de mantenimiento y el grado de aceptación del usuario hacia dicho sistema, producto o programa. Para evaluar el grado de usabilidad de las interfaces y detectar precozmente el mayor número de problemas se diseñan planes de evaluación que seleccionarán y combinarán distintas herramientas teniendo en cuenta las características de la interfaz, los objetivos prioritarios de la evaluación y la fase de desarrollo del producto o prototipo objeto del análisis; así como cualesquiera otras circunstancias o condicionantes, sea el coste o el tiempo, la formación del personal, etc.

La Evaluación Heurística es una herramienta de evaluación de la usabilidad basada en la intervención de expertos. Consiste en la comprobación de los elementos de una lista de heurísticas que, por su naturaleza, no pueden comprobarse automáticamente y requieren del criterio de un experto. Supone un método de bajo coste y buen rendimiento de detección de fallos que se adapta a distintos productos y objetivos, a las características de las distintas fases de desarrollo y, en definitiva, a muy distintos planes de evaluación.

En la literatura coexisten distintas propuestas de listas heurísticas de carácter general que, en dominios específicos, deben adaptarse a las características del contexto para maximizar su eficacia. Un ejemplo es la heurística propuesta en esta Tesis, adaptada a la interacción con interfaces móviles. Estas listas adaptadas también se han multiplicado en la literatura atendiendo a distintos aspectos de los planes de evaluación. De este modo el volumen de propuestas a disposición de los evaluadores, para su selección y potencial combinación, resulta bastante inmanejable.

Tomando un contexto tan específico como los videojuegos —juegos a partir de ahora—, sin embargo, una revisión sistemática de la literatura permite observar que la lista generalista de Nielsen sigue siendo la más usada, incluso cuando resulta insuficiente a la hora de abarcar las características específicas y novedosas de este dominio de aplicación.

Partiendo de las hipótesis de que el gran número de propuestas, con distintas características y objetivos, supone un obstáculo a la hora de que los evaluadores puedan seleccionar la más ajustada a su caso y de que incluso una selección ajustada desaprovecharía potenciales ventajas de otras propuestas, se ha desarro-

llado una herramienta denominada MUSE (*Meta-heUristics uSability Evaluation tool*) que es capaz de generar una lista heurística personalizada para un plan de evaluación específico tomando como base las distintas propuestas de la bibliografía y automatizando al máximo su selección y filtrado.

Una evaluación preliminar de la herramienta aventura resultados prometedores en cuanto a su capacidad de mejorar la detección de vulneraciones de la usabilidad frente al uso de herramientas genéricas en el caso del contexto específico de los juegos y, particularmente, cuando los evaluadores involucrados no son expertos.

La presente Tesis comprende, por tanto, tres trabajos principales:

- La construcción y análisis de una heurística adaptada a un contexto específico: interfaces móviles.
- Una revisión sistemática de la bibliografía que presenta el estado de las evaluaciones de usabilidad de juegos en el ámbito académico.
- El desarrollo de la herramienta MUSE usando como base la concepción modular de las heurísticas y su manejo desde una metaheurística capaz de adaptarse, potencialmente, a nuevos contextos específicos y características concretas de los planes de evaluación, generando listas más eficaces.

# Índice general

|                                                                                                         |            |
|---------------------------------------------------------------------------------------------------------|------------|
| <b>Agradecimientos .....</b>                                                                            | <b>I</b>   |
| <b>Resumen .....</b>                                                                                    | <b>III</b> |
| <b>PARTE I: MEMORIA.....</b>                                                                            | <b>1</b>   |
| <b>Introducción.....</b>                                                                                | <b>3</b>   |
| Bloque 1: Evaluación Heurística de interfaces móviles .....                                             | 10         |
| Bloque 2: Revisión sistemática de la evaluación de la usabilidad de juegos en el entorno académico..... | 11         |
| Bloque 3: Evaluación Heurística de juegos: un enfoque modular.....                                      | 15         |
| <b>Objetivos .....</b>                                                                                  | <b>19</b>  |
| <b>Resumen de resultados .....</b>                                                                      | <b>21</b>  |
| Bloque 1: Evaluación Heurística de interfaces móviles .....                                             | 21         |
| Bloque 2: Revisión sistemática de la evaluación de la usabilidad de juegos en el entorno académico..... | 24         |
| Bloque 3: Evaluación Heurística de juegos: un enfoque modular.....                                      | 31         |
| <b>Conclusiones .....</b>                                                                               | <b>37</b>  |
| <b>Bibliografía.....</b>                                                                                | <b>41</b>  |
| <b>PARTE II: PUBLICACIONES .....</b>                                                                    | <b>45</b>  |
| <b>Heuristic Evaluation on Mobile Interfaces: A New Checklist.....</b>                                  | <b>47</b>  |
| <b>Academic methods for usability evaluation of serious games: a systematic review .....</b>            | <b>49</b>  |
| <b>Heuristic evaluation on games: a modular approach.....</b>                                           | <b>51</b>  |



# Índice de figuras

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| Figura 1_Métodos de Evaluación Heurística de la usabilidad .....                                | 5  |
| Figura 2_Restricciones específicas en entornos móviles .....                                    | 23 |
| Figura 3_Interfaces innovadoras utilizadas en un 42% de los estudios .....                      | 25 |
| Figura 4_Histograma de frecuencias del número de usuarios involucrados en las evaluaciones..... | 26 |
| Figura 5_Histograma de frecuencias del número de expertos involucrados en las evaluaciones..... | 26 |
| Figura 6_Técnicas aplicadas en las evaluaciones basadas en <i>playtesting</i> .....             | 27 |
| Figura 7_Técnicas aplicadas en los análisis posteriores al uso .....                            | 27 |
| Figura 8_Técnicas aplicadas en las evaluaciones basadas en expertos .....                       | 28 |
| Figura 9_Listas heurísticas utilizadas en las evaluaciones heurísticas.....                     | 29 |
| Figura 10_Nube de metadatos .....                                                               | 33 |



# Índice de tablas

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| Tabla 1_ Criterios de inclusión y exclusión para la búsqueda sistemática ..... | 24 |
|--------------------------------------------------------------------------------|----|









**Parte I**

**Memoria**



# Capítulo 1

## Introducción

*Good design is obvious.  
Great design is transparent.  
Joe Sparano*

Un sistema de información es un conjunto de componentes hardware y software diseñados para dar de forma conjunta y colaborativa respuesta a una necesidad u objetivo funcional. Precisamente, una de las primeras tareas necesarias para su diseño es describir formal y exhaustivamente ese objetivo funcional descomponiéndolo en los llamados requisitos del sistema. La lista de requisitos que cubre un sistema se elabora mediante la descripción detallada de las necesidades atómicas que en conjunto conducen a la cobertura de la necesidad global a la que se quiere dar respuesta. Estas necesidades pueden ser funcionales —“el sistema debe hacer algo”— tratándose entonces de los llamados requisitos funcionales o bien pueden referir a la necesidad de cubrir la funcionalidad de una forma determinada —“cómo el sistema debe hacer ese algo”— tratándose, en este caso, de los llamados requisitos no funcionales. Los requisitos no funcionales (*non-functional requirements, NFR*) [1] comprenden aspectos generales a cualquier sistema como la seguridad en cualquiera de sus dimensiones [2] —confidencialidad, integridad, disponibilidad, autenticidad, trazabilidad—, la capacidad para atender las peticiones que va a recibir el sistema a lo largo el tiempo —disponibilidad, capacidad y continuidad—, la facilidad de mantenimiento y evolución, la capacidad de interoperar con otros sistemas, la confiabilidad, etc., y, adicionalmente, cuestiones particulares en dominios específicos como pueden ser la precisión numérica [1].

Ambos grupos de requisitos deben ser tenidos en cuenta desde las primeras fases del ciclo de vida. Sin embargo, los requisitos no funcionales son tradicionalmente pospuestos en el ámbito de la ingeniería y menos atendidos en los procesos de desarrollo [1]. Este retraso en su atención conlleva una merma de la calidad final de los productos y un aumento de los costes de fabricación y mantenimiento. Además, suponen un criterio diferenciador de orden comercial ya que a menudo se utilizan como criterio de selección a la hora de seleccionar un producto de entre aquellos que cubren la misma funcionalidad.

Entre los requisitos no funcionales generales a cualquier sistema centraremos este trabajo en la usabilidad. La usabilidad se define como el grado en el que un producto puede ser usado por usuarios específicos para conseguir unos objetivos determinados con eficacia, eficiencia y satisfacción en un contexto dado de uso [3]. Todos los sistemas exigen cierto grado de usabilidad para poder ser operados. Además, la atención a este requisito es fundamental en términos de

coste y tiempo ya que se estima que el 80% de los errores detectados en los sistemas de información provienen de la interfaz de usuario, de los cuales, alrededor de un 64% son causados por problemas de usabilidad [4]. Esto supone que la desatención a la usabilidad como requisito no funcional pueda llegar a ser el origen de la mayoría de desarrollos correctivos que necesite un sistema de información a lo largo de su ciclo de vida.

De la anterior definición de usabilidad podemos extraer otro concepto íntimamente relacionado: la accesibilidad. Más adelante ahondaremos en la relación entre ambos conceptos pero podemos decir que cuando los usuarios específicos a los que se refiere la definición de usabilidad son usuarios con capacidades motoras o sensoriales (ya sean visuales, auditivas, etc.) especiales, temporal o permanentemente, hablamos de accesibilidad. Formalmente, la accesibilidad es el grado en el que productos, sistemas, servicios, entornos e instalaciones pueden ser usados por un conjunto de personas con el más amplio rango de características y capacidades para conseguir un objetivo concreto en un contexto específico de uso [5]. Si bien el objetivo ideal de la accesibilidad es la cobertura universal, la práctica obliga a muestrear quiénes serán los usuarios potenciales del sistema y cuáles serán sus características. Cómo de amplia sea la cobertura del espectro de usuarios finales potenciales depende de objetivos comerciales como la relación coste-beneficio y también de cuestiones legales. De hecho, si enfrentáramos la usabilidad y la accesibilidad como conceptos, ésta última ha sido tradicionalmente mejor atendida en su condición de requisito no funcional debido a que en muchos países existen exigencias legales que obligan al cumplimiento de determinados niveles de accesibilidad bien para sistemas del ámbito de la administración pública o bien en el de la industria privada o en ambos. Además, veremos que la evaluación de la accesibilidad es más automatizable por naturaleza que la de la usabilidad, lo que contribuye a que sea más fácil y económico considerarla en los desarrollos.

En realidad, los límites entre usabilidad y accesibilidad están muy difusos. Una aproximación que en nuestra experiencia ha resultado ser útil en la práctica es considerar que la accesibilidad está enfocada a permitir el uso de los sistemas —de forma eficaz, eficiente y satisfactoria— a usuarios con discapacidades sensoriales o motrices mientras que la usabilidad lo haría atendiendo a las capacidades cognitivas de los usuarios. No existe consenso en la bibliografía en cuanto a esta distinción pero la expuesta es una primera aproximación que, en general, puede resultar muy útil en la práctica.

Otra aproximación posible sería considerar que la atención a la accesibilidad se engloba dentro del paraguas de la atención a la usabilidad. Así, todo sistema usable sería accesible aunque no todo sistema accesible sería usable. Esto sólo obliga a añadir a la definición de usabilidad expuesta anteriormente un hincapié específico en la inclusión de usuarios con capacidades diversas. Como hemos di-

cho, más allá del ideal de la inclusión universal, el grupo de usuarios objetivo y su variabilidad vendrá condicionado por otras circunstancias del sistema, su proceso de fabricación, sus objetivos comerciales y su marco legal.

En cualquier caso, centrándonos en la usabilidad y volviendo de nuevo a su definición, de ella se obtiene que el uso de un sistema de información como mediador o herramienta para conseguir unos objetivos tiene que producirse con eficacia, eficiencia y satisfacción. Así, la eficacia se refiere a la capacidad de los usuarios para completar tareas que les acercan a los objetivos que quieren cumplir. La eficiencia hace referencia a la inversión de recursos —generalmente bajo los criterios de tiempo y coste— que los usuarios tienen que hacer en el uso del sistema y que, en el mejor de los casos, será la mínima imprescindible. Y la satisfacción gira en torno a las expectativas, impresiones y sensaciones finales que los usuarios tienen sobre la facilidad de uso, confiabilidad, utilidad, etc. del sistema de información. Estas impresiones son en muchos casos determinantes en el éxito de los productos, la difusión de su uso, su comercialización y en el grado de adhesión de los usuarios.

A la hora de medir la usabilidad bajo estos tres aspectos, una primera aproximación tradicional es obtener medidas cuantitativas basadas en métricas como, por ejemplo, el tiempo empleado en una tarea, el número de veces que se toman decisiones o caminos no óptimos o, incluso, el tiempo en el que la mirada del usuario permanece sobre una determinada área de la interfaz. Todos estos valores son cuantificables y su recogida automatizable pero una valoración completa de todos los aspectos de la usabilidad exige tener en cuenta también información cualitativa relacionada con aspectos subjetivos o informales, cuya recogida y tratamiento es difícilmente automatizable. El manejar e incluir este tipo de información supone un desafío a la hora de definir metodologías e integrar los procesos de evaluación en el ciclo de vida de desarrollo de los sistemas.

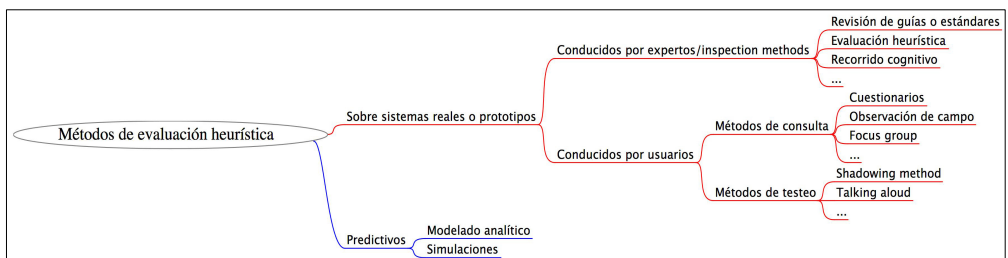


Figura 1\_Métodos de Evaluación Heurística de la usabilidad

Los métodos de evaluación de la usabilidad se pueden clasificar en función de las características del producto sobre el que se aplican [6, 7, 8, 25, 9]. Se muestra un resumen en la Figura 1. Así, encontramos métodos orientados a sistemas en producción o prototipos y por otra parte, métodos orientados a evaluar

diseños formales de sistemas aún sin desarrollar. Para la evaluación de diseños formales se utilizan métodos con una fuerte base matemática que, a su vez, se dividen en métodos basados en simuladores o basados en modelado analítico; ambos tratarían de predecir la usabilidad del producto al que dan lugar las especificaciones analizadas.

Por otra parte, y con mayor difusión fuera del ámbito académico encontramos los métodos orientados a evaluar sistemas o prototipos reales. A su vez, estos se clasifican en base a los perfiles de los participantes y conductores de la evaluación: así tenemos los métodos en los que fundamentalmente se evalúa la interacción real de un grupo de usuarios con el sistema y por otra parte, los métodos en los que en la evaluación sólo intervienen usuarios expertos.

Las evaluaciones conducidas por expertos se conocen también como *métodos de inspección*. Estos usuarios expertos lo son en el dominio del problema o, preferiblemente, en la propia disciplina de análisis y evaluación de la usabilidad; generalmente, es conveniente que sean independientes del equipo de desarrollo o propietarios del sistema. Entre estos métodos se encuentran las listas de comprobación de estándares o buenas prácticas que se basan en la verificación del cumplimiento de cada uno de los elementos de la lista. Cuando el resultado del chequeo está sujeto a la subjetividad del evaluador y por su propia naturaleza no puede automatizarse, se trata del método de Evaluación Heurística.

En cualquier caso, los métodos de evaluación mediante expertos conducen a conclusiones limitadas. Es ampliamente discutido en la bibliografía el enfrentamiento en cuanto a bondad de estos métodos respecto a aquellos en los que se involucra a usuarios y se observan interacciones reales [10, 11, 12]. Sin embargo, las evaluaciones conducidas por expertos requieren menor tiempo y coste ya que se evita el reclutamiento de usuarios, la preparación de las instalaciones, el análisis de un mayor volumen de datos, etc. y por tanto, la relación coste-beneficio es conveniente en muchos casos[25]. Un plan de evaluación óptimo debería combinar distintos tipos de evaluaciones para reforzar debilidades y aprovechar las ventajas de cada herramienta de evaluación.

Además, la evaluación con usuarios, por su naturaleza interactiva y por su coste, suele destinarse a las fases finales de desarrollo, cuando los prototipos son ya de alta fidelidad, muy cercanos a la versión final del sistema. Estos métodos pueden basarse en entrevistas con los usuarios con distinto grado de formalización según el método; se pueden pasar cuestionarios, observar a los usuarios mientras usan normalmente el sistema como en el *shadowing method* o concertar sesiones de grupo que se dirigen de forma metodológica en los llamados *focus groups*. Por otra parte, es habitual presentar a los usuarios un prototipo del sistema, una versión final o incluso dos versiones que difieren en algún aspecto de su in-



teracción, estos son los *testing methods*. Se recoge información de la interacción de los usuarios con el sistema presentado y se analiza combinando métricas para los valores cuantitativos con un análisis cualitativo de la información no cuantificable.

El conjunto de los métodos propuestos dentro de las categorías anteriores conforma una batería de herramientas que están a disposición de los evaluadores. Cada una de ellas tiene ventajas y beneficios y se adecúan a determinados productos y circunstancias del análisis. En el plan de evaluación deberían, como apuntábamos anteriormente, seleccionarse y combinarse los distintos métodos adaptando dicho plan a las características del sistema objetivo, a sus usuarios y a las condiciones de producción y distribución y, también, a factores como el conocimiento previo de los equipos de evaluación respecto de las metodologías a aplicar, el coste permitido, la planificación de tiempos, etc.

Por su parte, la Evaluación Heurística es un método de evaluación basado en expertos que presenta diversas ventajas y que se adapta muy bien a diversos productos y planes de evaluación. Como decíamos, se basa en la comprobación de ítems de una lista cuya verificación no es automatizable. El término heurística hace referencia al componente subjetivo que impide la automatización. Los resultados obtenidos están fuertemente influenciados por la experiencia de los expertos y su capacidad para interpretar los enunciados. Sin embargo, se han propuesto metodologías de aplicación que permiten minimizar la variabilidad de resultados debido al sesgo de los expertos [13].

La Evaluación Heurística es un método de bajo coste frente a los métodos que involucran a usuarios, se adapta muy bien a todas las fases de desarrollo y a los subproductos parciales que éstas van generando y detecta la mayoría de errores de usabilidad de grano grueso. Se estima que un 80% [14] con un número pequeño de evaluadores, de tres a cinco [15, 49], que a su vez pueden tener distintos niveles de conocimiento sobre usabilidad [16, 18]. De entre los métodos basados en expertos es el de uso más frecuente [17]. Por otra parte, y como con todos los métodos basados en expertos, conviene complementar el plan de evaluación con otros métodos y, siempre que sea posible, con evaluaciones en las que se involucre a los usuarios.

Es importante hacer hincapié en que la mayoría de problemas detectados mediante Evaluación Heurística suele ser de gran calado, es decir, de grano grueso, mientras que los detectados por métodos que involucran a usuarios suelen estar relacionados con cuestiones de detalle, de grano fino [18]. La corrección de problemas graves es muy costosa y tanto más conforme avanza el ciclo de vida del sistema. Es por ello que deben ser eliminados lo antes posible. Recordemos que se estima que el 80% de los errores detectados en los sistemas de información provienen de la interfaz de usuario y de ellos un 64% son causados problemas de

usabilidad. La influencia de la usabilidad en el coste total en mantenimientos correctivos es importante y debe ser tenida muy en cuenta.

Sobre las listas de comprobación usadas en la Evaluación Heurística, existen numerosas propuestas en la bibliografía. La lista de mayor popularidad[17] es la enunciada por Jacob Nielsen en 1990[18] e incluye diez elementos de evaluación:

1. Visibilidad del estado del sistema. Los usuarios deben poder conocer siempre qué está pasando en el sistema y en qué estado se encuentra para ello debe mostrárseles una retroalimentación adecuada y a tiempo.

2. Utilizar el lenguaje de los usuarios. El sistema debe usar un vocabulario y una forma de expresión que a los usuarios les resulte familiar y evitar términos técnicos o crípticos que no les faciliten información.

3. Control y libertad para el usuario. Los usuarios deben poder revertir sus acciones en el sistema con opciones de deshacer y rehacer para ofrecer al usuario la máxima libertad de interacción.

4. Consistencia y estándares. El sistema debe seguir las convenciones habituales en su dominio de aplicación y seguir los estándares habituales en sistemas de su mismo tipo. De esta forma se cumplirán las expectativas y mapas mentales de los usuarios aprovechando su conocimiento previo.

5. Prevención de errores. Es prioritaria la prevención de errores frente a cualquier otra medida como informar adecuadamente del error. Para evitar errores hay que contener las posibles acciones accidentales de los usuarios y establecer confirmaciones cuando las acciones tengan consecuencias irreversibles.

6. Minimizar la carga de la memoria del usuario. El sistema no debe apoyarse en la memoria del usuario y debe mostrarle siempre toda la información que necesita a la hora de tomar cualquier decisión. Lo mismo aplica a las instrucciones de uso que deben ser innecesarias o fácilmente accesibles durante el uso del sistema.

7. Flexibilidad y eficiencia de uso. El sistema debe adaptarse al uso por usuarios con distintos niveles de conocimiento. Así, proporcionar atajos de teclado o la personalización de acciones frecuentes puede ayudar a los usuarios más avanzados en el uso más rápido y cómodo del sistema.

8. Diálogos estéticos y diseño minimalista. Se debe evitar el cargar la interfaz con información irrelevante ya que supone distracción para el usuario y aumenta los requerimientos de atención hacia éste.

9. Ayudar a los usuarios a reconocer, diagnosticar y recuperarse de los errores. Los mensajes de error deben mostrarse en el lenguaje del usuario y ser suficientemente significativos.

10. Ayuda y documentación. Se debe diseñar el sistema de modo que la consulta a la ayuda y la documentación sea innecesaria pero en todo caso, debe accederse fácilmente a la información, ser contextual y concisa.

Los diez elementos de la lista heurística de Nielsen son lo suficientemente genéricos como para poder aplicarlos a cualquier sistema y tan completos como para conseguir excelentes resultados a la hora de detectar errores de usabilidad de grano grueso. Su popularidad la lleva a ser no sólo la más usada sino a ser la base de muchas de las nuevas listas heurísticas propuestas en la bibliografía.

Sin embargo, existen sistemas con características específicas, bien debidas al tipo de usuario a que se destinan, bien por los modelos de interacción que plantean o bien por sus condiciones de uso, que crean contextos específicos que no están cubiertos por las listas heurísticas generalistas. El propio Nielsen ya declaró que era necesario adaptar las heurísticas cuando los dominios de aplicación de los sistemas a evaluar fueran suficientemente específicos[19]. Así, además de nuevas propuestas de evaluación generalistas, se han publicado numerosas propuestas de herramientas de Evaluación Heurística orientadas a cubrir los contextos específicos de ciertos sistemas. Por ejemplo, existen herramientas que tienen en cuenta los condicionantes de uso de las plataformas móviles[8] o bien heurísticas destinadas a analizar la usabilidad de pantallas ambientales que muestran información efímera en el entorno del usuario [20].

En la presente Tesis abordamos el problema de elaboración de heurísticas para contextos específicos.

Como trabajo inicial se acometió la confección metodológica y evaluación de una lista de Evaluación Heurística centrada en entornos móviles que se aplicó en el proyecto de investigación Procur@[21]. Posteriormente, dentro del mismo proyecto surgió la necesidad de evaluar la usabilidad de otro entorno específico, los juegos. Se decidió entonces revisar la bibliografía de forma sistemática. A partir de esta revisión se obtuvo una panorámica del estado del arte en cuanto a evaluación y aseguramiento de la usabilidad en juegos. Teniendo presentes las ventajas de las evaluaciones conducidas por expertos y la Evaluación Heurística para fases preliminares de desarrollo, estado en que se encontraba la herramienta desarrollada en Procur@, se centró el interés en esta técnica. En la revisión se había detectado que si bien la heurística de Nielsen y otras de uso generalista son conocidas por los desarrolladores y ampliamente aplicadas en la evaluación de la usabilidad de juegos, otras propuestas más recientes y centradas en las particularidades de los juegos no son usadas por los desarrolladores a pesar de sus ventajas[17]. Por ello pareciera que confeccionar una nueva lista, como la desarrollada para entornos móviles, podría conducir a una difusión muy restringida de su uso.

Establecimos como hipótesis que este problema de adopción de las nuevas propuestas, potencialmente ampliable a otros contextos específicos, pudiera deberse a la dificultad que supone para los desarrolladores conocer todas las herramientas disponibles, así como discriminar de entre ellas las aplicables para un producto y contexto de evaluación dado. Ante este problema, se conceptualiza una concepción modular de las listas heurísticas aplicables para la evaluación de la usabilidad de juegos que permite aunar los esfuerzos de los distintos autores y las propuestas ya existentes facilitando el acceso y uso a los desarrolladores. Esta propuesta puede entenderse como una meta-heurística capaz de aglutinar propuestas actuales y dejar abierta la posibilidad de inclusión de propuestas futuras. Además, esta idea podría, en un futuro, aplicarse a otros contextos específicos más allá de los juegos.

Para gestionar la meta-heurística y evaluar sus posibilidades se desarrolla, además, una herramienta que a partir de las heurísticas propuestas en la bibliografía, genera listas personalizadas escogiendo los elementos de evaluación y los módulos más convenientes para cada plan de evaluación. En los siguientes apartados abordaremos con más detalle esta propuesta.

## **Bloque 1: Evaluación Heurística de interfaces móviles**

La presente Tesis se inició en el período en que la doctoranda trabajó, como investigadora contratada, en el marco del proyecto nacional Procur@: Plataforma abierta de soporte a la prevención y rehabilitación de enfermedades neurodegenerativas (IPT-2011-1038-900000) en el que participaron el Centro Tecnológico del Juguete-AIJU, la Universidad de Sevilla, la Universidad de Zaragoza, las empresas Logica, Solutio, Flowlab y el Hospital Universitario Virgen del Rocío. Se construyó una plataforma tipo red social que atendía a distintas necesidades de colectivos de pacientes enfermos de Parkinson, Alzheimer y accidente cerebrovascular[21]. Estas necesidades comprendían cuestiones emocionales de mantenimiento y robustecimiento de sus redes sociales, la mejora de las comunicaciones entre pares y con los servicios asistenciales y la provisión de herramientas terapéuticas basadas fundamentalmente en juegos serios que los pacientes podían practicar desde casa con la adecuada monitorización sanitaria. Se habla de juegos serios cuando el juego tiene alguna aplicación terapéutica, docente, de refuerzo de la actividad física o de cualquier otro aspecto que complementa al de la mera diversión [26]. Para el acceso a todas estas funcionalidades se diseñó un acceso desde equipos de escritorio pero también desde dispositivos móviles. En esta segunda interfaz, destinada especialmente a los usuarios finales, surge la necesidad de extremar el cuidado en el aseguramiento de la usabilidad.

Dado el perfil de los usuarios de Procur@ resultaba crítico para el éxito del proyecto el aseguramiento de unos niveles adecuados de usabilidad y se estableció un plan de evaluación que contemplaba tanto la evaluación por expertos como intervenciones con usuarios que en distintas fases se basaron en *focus groups* y observación en pruebas de testing basadas en *thinking aloud*.

En el apartado de evaluación por expertos se eligió la Evaluación Heurística por las ventajas anteriormente comentadas pero se descubrió que la aplicación de heurísticas genéricas no conseguía alcanzar el nivel de detalle requerido. Con estas necesidades se inició la construcción y análisis de una heurística específica para el contexto móvil.

Más allá del contexto del proyecto Procur@, la emergencia y difusión de los dispositivos móviles, sean teléfonos, PDAs o sistemas híbridos, ha creado nuevos modelos de interacción que suponen nuevos retos para la evaluación de la usabilidad. En concreto, en el momento de publicación de nuestro trabajo parecía no existir ninguna herramienta de evaluación de alto detalle que permitiera evaluar mediante heurísticas el diseño inicial de una interfaz móvil.

Tras un proceso sistemático de elaboración se obtuvo una lista de 230 subheurísticas clasificadas bajo 13 heurísticas cuya evaluación preliminar apuntó a que podía ser una herramienta ventajosa [22] Esta lista heurística, descrita en el primer artículo que forma parte del compendio que constituye la presente Tesis, ha sido referenciada por numerosos trabajos académicos que la han utilizado para evaluar interfaces móviles. Sin embargo, como decíamos anteriormente, mediante un análisis sistemático de la bibliografía, detectamos que la difusión de este tipo de propuestas de Evaluación Heurística orientadas a entornos específicos era bastante restringida frente al uso de listas generalistas[17]. Este hecho dirige los siguientes trabajos de la presente Tesis.

## **Bloque 2: Revisión sistemática de la evaluación de la usabilidad de juegos en el entorno académico**

Los juegos contienen una interactividad esencial que los convierte en un caso de estudio especialmente relevante en el ámbito de la interacción hombre-máquina. Esto es aún más notorio en el caso de los juegos serios; estos se definen como juegos cuyo propósito está más allá de la diversión[26]. Son juegos orientados a fines pedagógicos, de entrenamiento, de cuidados médicos o terapéuticos, entre otros, y son utilizados por grupos de usuarios que, en muchos casos, requieren de una interacción adaptada a sus características especiales sean sensoriales, motoras o cognitivas. En muchos casos, estos grupos de usuarios son niños, an-

cianos, personas con movilidad reducida, etc. lo que convierte la usabilidad y la interacción en un aspecto aún más importante a la hora de determinar el éxito de adopción y funcionamiento de dichos juegos. Además, mientras que los usuarios habituales de juegos comparten un conocimiento acerca de convenciones de interacción y modelos mentales comunes, los usuarios de juegos serios son en muchos casos personas que no han tenido relación anterior con juegos y que, además de una resistencia inicial, necesitan que la interacción sea especialmente amigable.

A la hora de evaluar la usabilidad de los juegos en general y los serios en particular, hay tres aproximaciones posibles: utilizar herramientas estándar asumiendo que las diferencias entre los juegos y el software generalista no son críticas; descartar cualquier herramienta previa y crear nuevos métodos de evaluación [15, 23, 31, 34, 35, 41, 46, 54, 55] especialmente adaptados; o adaptar herramientas existentes de enfoque generalista, sin descartarlas por completo, al nuevo ámbito de aplicación [17] y, por tanto, a las características especiales de los juegos. No parece haber acuerdo sobre qué enfoque es el más ventajoso.

Es imprescindible ahondar, de partida, en cuáles son las características de los juegos que los diferencian del software generalista o sistemas de productividad para los que las herramientas de usabilidad han sido concebidas tradicionalmente. Por una parte, el diseño de los sistemas de productividad debe ayudar a que sea fácil aprender a usarlos, que el manejo sea sencillo y que se pueda alcanzar de forma rápida un nivel experto de uso. Sin embargo, el fin último de los juegos no es utilitario, no son herramientas que persigan el completar una tarea, sino que se usan por mera diversión o, en el caso de los juegos serios, sus fines se consiguen aprovechando la diversión. Si se puede aprender a manejar un juego y alcanzar el nivel experto muy rápido no captará el interés de los usuarios. En el diseño de los juegos no sólo se permite, sino que es deseable, mantener ciertos obstáculos para el uso y el aprendizaje que supongan retos para los jugadores [23, 24]. Pero estos obstáculos deben estar cuidadosamente seleccionados y diseñados y no deben ser nunca causados por problemas de usabilidad. De hecho los fallos de usabilidad son especialmente problemáticos en el caso de los juegos precisamente porque los usuarios juegan por diversión y son libres de abandonar el juego cuando les resulte fastidioso o aburrido. Así que el objetivo final debería ser reducir los obstáculos que impiden la diversión pero no los obstáculos que hay que salvar para alcanzar logros. La interacción de los jugadores con la interfaz, la narrativa y cualquier otro aspecto del juego deben ser especialmente fluidos, sin problemas de usabilidad u obstáculos distintos de los considerados por diseño, ni interrupciones que distraigan al jugador y puedan afectar a sus sensaciones de satisfacción [25].

Por otra parte, la usabilidad en juegos se ha definido como el grado en el que un jugador es capaz de aprender, controlar y entender un juego [31]. La curva de aprendizaje de un nuevo usuario enfrentado a un juego debe ser cuidadosa-

mente optimizada. Como lema fundamental se suele decir que un buen juego es “fácil de aprender y difícil de dominar”[23, 24] de manera que un nuevo jugador debe ser capaz de entender y realizar las acciones básicas del juego con mucha rapidez y facilidad y debe exigírsele mayor atención a la hora de explorar formas óptimas de alcanzar los logros definidos en el juego.

Decíamos que el objetivo último de un juego es la diversión. Es necesario matizar esta afirmación. Con el surgimiento y popularización de los juegos serios, los objetivos se amplían. De hecho, los juegos serios se definen por su propia condición de tener algún propósito más allá de la mera diversión. Son usados con fines formativos, pedagógicos, de entrenamiento, de promoción de la salud, terapéuticos, etc.[26] El aseguramiento de estos propósitos particulares requiere una evaluación específica que podría considerarse parte de la batería de pruebas orientadas a asegurar el cumplimiento de los requisitos funcionales pero, además, los juegos serios siguen manteniendo como objetivo la diversión ya que ésta favorece la adhesión de los usuarios al entrenamiento o aprendizaje. Así, el carácter lúdico de la interacción sigue siendo característico y evaluable. Por tanto el diseño y evaluación de su usabilidad deberá atender a este aspecto específico.

Además, en ocasiones, y especialmente en el caso de los juegos serios, los usuarios objetivo pueden tener necesidades especiales. Muchos juegos para rehabilitación o educación están orientados a niños o ancianos con características cognitivas concretas o bien a personas con discapacidades sensoriales y/o motrices que establecen límites a los mecanismos de interacción con el juego. Así pues el diseño de la curva de aprendizaje y los distintos modos de juego según el nivel de experiencia debe ser aún más cuidadoso en este tipo de casos.

Otro reto que plantea la evaluación de usabilidad de juegos es la aparición constante de innovaciones que pretenden atraer a los compradores y usuarios hacia los nuevos juegos que se presentan en el mercado. Estas innovaciones pueden consistir típicamente en la utilización de dispositivos hardware novedosos o en la aplicación de mejoras tecnológicas. Estas innovaciones conducen en muchos casos a nuevos modelos de interacción entre usuarios y sistema que plantean retos a la evaluación de la usabilidad ya que surgen cuestiones no atendidas tradicionalmente. Por ejemplo, la popularización de los juegos en red obliga a tener en cuenta, como parte integral de la usabilidad, el modo en que la comunicación de los jugadores entre sí interfiere en la fluidez del juego o cómo se integra esta comunicación en la interfaz de usuario.

Por último, los juegos se clasifican en géneros que determinan distintas formas de interacción típicas para cada uno de ellos y que deben tenerse en cuenta a la hora de asegurar y evaluar la usabilidad.

Una vez descritas las características particulares que se manejan a la hora de evaluar la usabilidad de los juegos será necesario reajustar el alcance y contenido de los tres aspectos cubiertos por la propia definición de usabilidad, esto es[3]: eficacia, eficiencia y satisfacción.

En cuanto a la eficacia, hemos visto que el objetivo último del juego no es apoyar a la realización de una tarea. Sin embargo, sí que podríamos seguir hablando de tareas a realizar si redefinimos el término: si bien las tareas en los sistemas de productividad son acciones que ayudan a los usuarios a alcanzar el objetivo de realización de un trabajo y que se van realizando en un orden determinado, en el caso de los juegos las tareas pueden ser consideradas acciones que producen un avance en el juego y que se ejecutan de forma azarosa; esto es típico del juego anárquico pero también ocurre en el juego basado en reglas[27][17], o incluso aunque el juego se desarrolle dentro del paradigma de ganar-perder. Mientras que algunos métodos de usabilidad requieren la descripción en detalle de las tareas[28][29][30][17], en el caso de los juegos esto no será posible debido a ese componente azaroso y de descubrimiento. En concreto el método de Evaluación Heurística no requiere una descripción detallada de las tareas pero sí puede hacerse referencia al concepto desde las heurísticas listadas para evaluar. Así pues, habrá que tener en cuenta esta ampliación de la definición del concepto de tarea.

En cuanto a la eficiencia, la inversión en recursos se puede considerar óptima tomando distintos criterios. Mientras que los errores son inadmisibles en el caso del software de productividad, para los juegos se consideran parte esencial de su uso. Muchas veces los errores son necesarios para la adquisición de habilidades que permiten finalmente conseguir los objetivos del juego[31]. Además, la tolerancia a los errores es aún mayor en el caso de los juegos serios ya que sus usuarios pueden ser propensos a cometerlos y porque el error puede formar parte de la terapia o el aprendizaje objetivo. Así, el concepto de error también debe ser redefinido para excluir acciones no especialmente eficaces, como pueden ser las exploraciones erráticas del entorno en juegos con narrativa. Hay que distinguir entonces los errores y dudas que surgen de un mal diseño de aquellos que forman parte de la naturaleza de exploración y descubrimiento del juego [32].

El tercer aspecto, la satisfacción, es hacia el que se han orientado el mayor número de iniciativas de evaluación adaptadas a los juegos [33, 23, 34, 35, 36]. Sin embargo, hay que hacer hincapié en que el término satisfacción es, en sí, usado bajo cierta controversia. Algunos autores consideran que la usabilidad no debe atender a cuestiones de estética o preferencias personales, de manera que en los juegos el entretenimiento, la adhesión y la narrativa, para estos autores, estarían fuera del ámbito de la usabilidad y se considerarían aspectos artísticos o técnicos[31]. Sin embargo, nosotros, por coherencia con la definición ISO de usabilidad, consideraremos que la satisfacción debe ser evaluada y considerada a la hora



de evaluar la usabilidad de un juego. En todo caso, las cuestiones relativas a aspectos motivacionales resultan difíciles de tratar ya que los procesos motivacionales se conocen menos que los cognitivos[36].

El concepto de satisfacción aplicado a juegos ha tomado en ocasiones el nombre de jugabilidad o experiencia de juego. De nuevo, es controvertido el uso de esta terminología dado que hay autores que consideran que la jugabilidad es meramente la forma que toma el concepto de usabilidad cuando aplica a los juegos y no uno sólo de los aspectos de su definición[37]. En la presente Tesis consideraremos que la jugabilidad o experiencia de juego atiende a la evaluación de la satisfacción de forma complementaria a la eficiencia y la eficacia.

Por si los márgenes de trabajo y de terminología no estuvieran suficientemente difusos, en la literatura han surgido nuevos conceptos como la experiencia de uso y la inmersión y el *flow*, particularmente importantes estos últimos en el caso de los juegos. La experiencia de uso es un concepto ampliado de la usabilidad que no sólo atiende a la interacción de uso del sistema sino a su adquisición y mantenimiento [38]. En cuanto a la inmersión y el *flow*, ambos términos se refieren a un uso óptimo en el que se alcanza un estado de concentración, de especial disfrute y total absorción mientras se está jugando[39]. Aunque la definición de estos términos y su conocimiento profundo está todavía en investigación [42] es importante atender a estos aspectos particulares cuando se evalúa la usabilidad. Por ejemplo, las interrupciones forzadas por las preguntas o intervenciones de los evaluadores durante las sesiones de testeo pueden alterar los resultados de la evaluación al acabar con la sensación de inmersión. Lo mismo aplicaría al requerimiento de verbalización por parte de los usuarios al aplicar la técnica *Think Aloud* que les exige explicar qué están haciendo, qué quieren hacer, qué están viendo en la interfaz... Por otra parte, la inmersión aumenta en los usuarios la capacidad de pasar por alto errores de usabilidad[40][41][42] y este prejuicio debe ser tenido en cuenta al procesar información subjetiva proporcionada por los usuarios o cuando es el propio experto quien ha alcanzado el estado de inmersión durante sus pruebas del juego.

### **Bloque 3: Evaluación Heurística de juegos: un enfoque modular**

Durante la realización del proyecto Procur@[21], anteriormente citado, surgió la necesidad de evaluar la usabilidad de juegos en estados preliminares de desarrollo. Pensando en usar Evaluación Heurística por sus ventajas de coste y su adaptación a prototipos iniciales se encontró, según la revisión sistemática de la

bibliografía[17], que se siguen usando listas de propósito general en lugar de heurísticas específicas. Además, en el contexto de los juegos el aseguramiento de la usabilidad es especialmente importante ya que hay quienes consideran que la principal característica de un juego es la interactividad[42][43].

En el apartado anterior ya matizamos el concepto de usabilidad al aplicarlo en juegos pero para el caso de la aplicación de Evaluación Heurística sobre ellos hay que destacar que el ciclo de vida de los juegos se caracteriza por ser producciones a gran escala, con gran inversión por parte de los fabricantes y que, por la alta competitividad empresarial, hay reticencia a la hora de exponer versiones parciales a usuarios que puedan filtrar información clave para conservar la originalidad de las nuevas propuestas [44]. Así, las evaluaciones conducidas por expertos parecen más ventajosas para las fases preliminares de desarrollo de manera que las evaluaciones con usuarios puedan posponerse. De entre las evaluaciones con expertos, la más usada es la Evaluación Heurística [17]. Se trata de un método adaptable a todas las fases de desarrollo que se convierte en una herramienta ideal para evaluar prototipos iniciales e intermedios o aspectos parciales de interacción que por novedosos pueda resultar de interés analizar específicamente. Este prototipo temprano y su evaluación permiten consolidar decisiones de diseño y ahorrar pérdidas cuantiosas de esfuerzo minimizando la posibilidad de fracaso final del producto.

Sin embargo, a pesar de la adaptabilidad de la Evaluación Heurística, las herramientas tradicionales están poco adaptadas al contexto específico de los juegos. Como ya dijimos, el propio Nielsen declaró la conveniencia de crear heurísticas para dominios específicos[19]. De este modo, surgen en la literatura diversas propuestas de listas heurísticas centradas en juegos que suelen orientarse a la evaluación de alguno de los aspectos de la usabilidad más que a una evaluación del conjunto de la definición.

Una de las primeras propuestas de aplicación de Evaluación Heurística al aseguramiento de la usabilidad de juegos fue la de Clanton[33] que dividía los elementos heurísticos en tres bloques: interfaz, mecánica y jugabilidad. El bloque dedicado a interfaz podría entenderse como el dedicado a conceptos tradicionales de interacción hombre-máquina mientras que mecánica y jugabilidad son bloques que se orientan a aspectos novedosos de la interacción más cercanos a los conceptos de narrativa, personajes, inmersión y motivación. De esta forma se abre el camino a propuestas posteriores que se centran especialmente en estos últimos aspectos de evaluación [33, 23, 34,36].

Otras propuestas se centran en aspectos concretos de los juegos como la movilidad[35] o la interacción cuando el juego es multijugador [54]. Pero también se produce una explosión de propuestas por la propia naturaleza innovadora de

los juegos: constantemente aparecen nuevas plataformas, géneros o dispositivos que introducen novedades que en muchos casos retan a las herramientas de evaluación de usabilidad disponibles.

Esta explosión y variabilidad de propuestas hace que para los evaluadores sea difícil seleccionar qué subconjunto de herramientas se ajusta mejor una evaluación concreta. Quizá por ello se sigue aplicando la heurística de Nielsen, de enfoque general, en la mayoría de los casos[17]. Esta heurística no sólo no está adaptada a este contexto específico sino que en algunos casos contraviene la propia naturaleza de los juegos[45]. Siendo conservadores, se puede considerar que la heurística de Nielsen puede ser útil pero no suficiente a la hora de evaluar los juegos[34]. Y es que, no sólo se trata de filtrar las herramientas útiles sino de combinarlas para cubrir todos los aspectos de la usabilidad que se quieren evaluar y que probablemente sean atendidos con distinto alcance en cada herramienta.

Una característica de las heurísticas es que tanto en la confección de nuevas listas de elementos como en el uso final de las mismas en las evaluaciones, pueden ser entendidas como conjuntos modulares que permiten la fragmentación y combinación de partes que sean aplicables frente a las que no proceda utilizar [46, 35]. El problema es que la cantidad de información a manejar, tanto en origen, para extraer heurísticas útiles, como en la combinación de las mismas, puede llegar a ser inmanejable.

Podemos ampliar la concepción modular al conjunto de heurísticas propuestas por los distintos autores considerando que toda la literatura está compuesta por módulos propuestos cada uno de ellos por distintas fuentes. Un procesamiento automático de estas fuentes de información podría generar una metaheurística que englobaría todas las heurísticas publicadas y desde la que sería posible entresacar los módulos y elementos aplicables para un plan de evaluación dado. Esta propuesta es el núcleo central del tercer y último artículo del compendio que conforma la presente Tesis. Con esta aproximación se pretenden aliviar los problemas de selección de herramientas heurísticas que eran la hipotética explicación que barajábamos ante la falta de difusión de las listas heurísticas para entornos específicos frente a las generalistas.

Teniendo en cuenta todo esto, se desarrolla la herramienta MUSE[42] (*Meta-heUristics uSability Evaluation tool*) que automatiza el proceso de selección de módulos a partir de un conjunto amplio y ampliable de listas heurísticas extraídas de la literatura. Además de una construcción inicial de dicha herramienta, se aborda una evaluación previa de los resultados obtenidos al aplicar la herramienta en la detección de problemas de usabilidad en juegos.



## Capítulo 2

# Objetivos

*Pluralitas non est ponenda sine necessitate. Ockham*

Como se ha apuntado brevemente en la introducción la presente Tesis se divide en tres trabajos interrelacionados: por una parte la elaboración metodológica de una heurística para interfaces móviles; por otra, la revisión sistemática de la literatura sobre los diferentes métodos de evaluación de la usabilidad en un entorno específico como son los juegos; y, finalmente, la aplicación de un enfoque modular a la generación de listas heurísticas orientadas al análisis de la usabilidad de juegos de forma personalizada a cada caso de aplicación concreto. Para cada bloque podemos formalizar un conjunto de objetivos que se detallan a continuación.

### Bloque 1: Evaluación Heurística de interfaces móviles

1. Adaptar un método para generar heurísticas en contextos específicos aplicándolo al desarrollo de una heurística centrada en entornos móviles
2. Describir formalmente el contexto de aplicación y el alcance del problema.
3. Seleccionar el conjunto de fuentes bibliográficas sobre las que aplicar el método anterior y obtener la lista de heurísticas tras su clasificación, reordenación y selección o descarte.
4. Evaluar de forma preliminar la utilidad de la lista heurística obtenida para la detección de problemas de usabilidad.

Bloque 2: Revisión sistemática de la evaluación de la usabilidad de juegos en el entorno académico. En este bloque, los objetivos se formalizan en base a las cuestiones (RQ-Research Questions) planteadas en la revisión sistemática.

5. Responder a las cuestiones:
  - (RQ1) ¿Cuál es la aproximación preferida al evaluar la usabilidad de juegos: descartar las técnicas tradicionales y desarrollar nuevos métodos; aplicar directamente herramientas tradicionales estándar bajo la hipótesis de que las diferencias entre juegos y software de productividad no son críticas; o adaptar las herramientas tradicionales existentes para superar dichas diferencias?

- (RQ1.1) ¿Se usan técnicas conducidas por tareas para evaluar juegos? ¿Se adaptan de alguna forma?
- (RQ1.2) ¿Se entiende que el concepto de rendimiento en juegos es distinto que en otros dominios? ¿Se redefine el concepto de error?
- (RQ1.3) ¿Las interrupciones del instructor se evitan conscientemente durante el test con usuarios? ¿Se utiliza la técnica *Think Aloud* de forma frecuente?
- (RQ1.4) ¿Se mide la satisfacción con una atención especial? ¿Se usan las nuevas propuestas de heurísticas para este objetivo?
- (RQ2) ¿Cuáles son las características más destacables de los juegos evaluados (por ejemplo, área de estudio) y del entorno de testeo (por ejemplo, número o edad de los usuarios, tipo de interfaz, etc.) en las evaluaciones que se describen en la literatura?

### Bloque 3: Evaluación Heurística de juegos: un enfoque modular

6. Desarrollar una herramienta capaz de proporcionar heurísticas personalizadas para un plan específico de evaluación que incluye los objetivos de evaluación, el estado de desarrollo del juego a evaluar y sus características específicas.
7. Presentar las heurísticas generadas por la herramienta en un nivel de abstracción lo suficientemente bajo como para ser utilizadas por evaluadores no expertos en usabilidad de manera que se facilite su uso por el mayor número de desarrolladores posible.
8. Verificar que la heurística personalizada generada por la herramienta debe conducir a mejores resultados que los obtenidos mediante el uso de listas heurísticas de propósito general.
9. Optimizar las listas de evaluación personalizada generadas por la herramienta minimizando el número de elementos no aplicables (falsos positivos) y la ausencia de elementos que pudieran ser útiles y no aparecen (falsos negativos) de entre el conjunto general de propuestas existentes en la bibliografía.

## Capítulo 3

### Resumen de resultados

*Ideas are worthless. Execution is everything.*  
Scott Adams

Describimos a continuación brevemente los resultados obtenidos en los tres bloques de trabajo. Inicialmente se genera una nueva propuesta de Evaluación Heurística para entornos móviles. Posteriormente se presenta la necesidad de acometer la evaluación de juegos. Previo a cualquier trabajo ejecutivo se inicia un estudio del estado del arte en forma de búsqueda sistemática para conocer cómo se está realizando la evaluación de usabilidad de juegos en la academia. Manteniendo el foco en la Evaluación Heurística y tras descubrir el problema de difusión de las nuevas propuestas, se crea una herramienta, desde una concepción metaheurística y modular, que permite la generación de heurísticas adaptadas aplicándola inicialmente a la evaluación de juegos.

#### Bloque 1: Evaluación Heurística de interfaces móviles

Para la construcción de la nueva propuesta de lista de Evaluación Heurística para móviles fue necesario formalizar distintas cuestiones del proceso. De entrada fue necesario adaptar alguno de los métodos propuestos en la bibliografía para la confección de nuevas heurísticas, en concreto el propuesto por Rusu et al. [47].

La metodología de construcción a seguir se definió en seis pasos (objetivo 1):

1. Definir claramente el alcance del problema y clasificar las características especiales de la interacción móvil
2. Reordenación de las heurísticas ya conocidas en una nueva compilación, entendiéndolas como chequeos generales que deben cumplirse para asegurar la usabilidad.
3. Después de construir una nueva clasificación de heurísticas, desarrollar una compilación de subheurísticas. Heurística se entenderá como el término que engloba una recomendación general mientras que subheurística hará referencia a elementos específicos a incluir en la lista. La principal di-

ferencia entre ambos conceptos estriba en el nivel de abstracción y en el nivel de experiencia requerido hacia el evaluador ya que los elementos más concretos están sujetos a una menor ambigüedad y requieren menor conocimiento de partida en el evaluador.

4. Enriquecer la lista de subheurísticas clasificadas bajo las heurísticas con elementos específicos del ámbito móvil extraídos de las evaluaciones publicadas en la literatura y en las guías de buenas prácticas de diseño.
5. Homogeneizar la redacción de los elementos y adaptar el lenguaje para evaluadores no expertos.
6. Evaluar preliminarmente la lista heurística obtenida.

En cuanto a la definición del contexto de aplicación y el alcance del problema (objetivo 2) se inició con la clasificación de los dispositivos móviles existentes:

1. *Feature phones*: Dispositivos básicos con pantallas pequeñas y teclados muy limitados que están adaptados especialmente a marcar números de teléfono
2. *Smartphones*: Teléfonos con pantallas de tamaño medio y teclados completos A-Z.
3. *Touch phones/Touch tablets*: Dispositivos con pantallas táctiles que cubren prácticamente todo el frontal.

Se centró el estudio en los *Touch phones/touch tables* al ser los dispositivos más populares en la actualidad y de uso por población general. Partiendo de la restricción del tipo de dispositivo se encontró que en la literatura se nombraban las restricciones a tener en cuenta en este tipo de dispositivos que se muestran en la Figura 2 y se describen brevemente a continuación:

- A) Capacidades limitadas en entrada/salida. Estas limitaciones están impuestas por los métodos de entrada, el tamaño de las pantallas y su resolución, el ancho de banda disponible así como la falta de confiabilidad de las redes inalámbricas.
- B) Movilidad y contexto cambiante. El uso de los dispositivos móviles se realiza en un entorno cambiante, con constantes distracciones. Este entorno cambiante es difícilmente reproducible en laboratorio y es un reto para la evaluación de usabilidad integrar esta información contextual.
- C) Tipos de tareas especialmente diferentes. Las nuevas tareas descritas en la literatura para los entornos móviles incluyen las búsquedas, bien para rescatar información, para realizar búsquedas más o menos específicas o para realizar comparaciones y selecciones. Otra tarea sería la navegación bien indirecta o condicionada, la visualización de vídeos, la lectura de noticias. Por otra parte existen tareas de comunicación como la comprobación del



correo electrónico o el acceso a redes sociales o foros. También hay que incluir la realización de transacciones comerciales, la disponibilidad de juegos y una actividad también a considerar: el uso para matar el tiempo.

- D) Acceso multidispositivo. Los modelos mentales en que se basa el diseño de aplicaciones y webs deben adaptarse a las distintas restricciones de interfaz en los distintos dispositivos.
  - E) Capacidad limitada de procesamiento y energía, incluyendo la duración de la batería, la conexión a la red que en muchos casos supone un coste añadido, latencias y memoria limitada.
  - F) Adopción. La adopción por parte de los usuarios está íntimamente relacionada con su percepción de la privacidad, la aceptación de la tecnología, el confort y la capacidad de personalización. Estos factores, relacionados íntimamente con la usabilidad, influyen notablemente en el éxito comercial de estos productos que han pasado por un proceso progresivo de adopción.
- G)

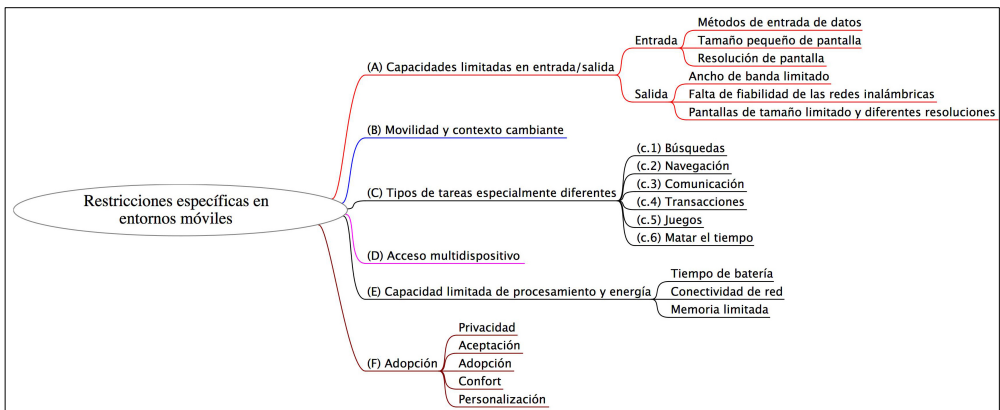


Figura 2\_Restricciones específicas en entornos móviles

Una vez delimitado el alcance del problema se seleccionaron fuentes bibliográficas para construir la clasificación en base a heurísticas (ver Figura 3), la incorporación de subheurísticas y el enriquecimiento de la lista con elementos específicos para plataformas móviles tomados de evaluaciones publicadas en la literatura y guías de buenas prácticas (objetivo 3). De este trabajo de recopilación surgen 83 heurísticas y 361 subheurísticas que una vez filtradas dan como resultado un conjunto de 13 heurísticas y 230 subheurísticas, 158 de ellas generales y 72 específicas para entornos móviles.

Una vez obtenida la lista anterior se realizó una evaluación inicial de la herramienta (objetivo 4). La herramienta se utilizó para refinar un prototipo de interfaz móvil desde tablet para el acceso a la red social construida en el proyecto Procur@[21].

## Bloque 2: Revisión sistemática de la evaluación de la usabilidad de juegos en el entorno académico

Los primeros pasos de la revisión sistemática se orientaron a formalizar los criterios de inclusión y exclusión (ver Tabla 1) para filtrar los artículos a analizar. De este modo se acabó considerando un conjunto de 187 estudios relevantes.

| Criterios de inclusión                                                                                                                                                                                                                                    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Se consideran los artículos que incluyen el término “usability” y “game” en el título, cuerpo o abstract</li> </ul>                                                                                              |
| <ul style="list-style-type: none"> <li>• Artículos publicados en revistas, conferencias (por ejemplo, simposios, <i>workshops</i>) y <i>magazines</i></li> </ul>                                                                                          |
| <ul style="list-style-type: none"> <li>• Artículos en inglés</li> </ul>                                                                                                                                                                                   |
| <ul style="list-style-type: none"> <li>• Cualquier fecha de publicación</li> </ul>                                                                                                                                                                        |
| <ul style="list-style-type: none"> <li>• Artículos que incluyen al menos una evaluación de usabilidad de un juego</li> </ul>                                                                                                                              |
| <ul style="list-style-type: none"> <li>• Si la evaluación está basada en la interacción con usuarios debe involucrar a más de 4 usuarios Si la evaluación se realiza mediante una evaluación conducida por expertos no se aplica ningún filtro</li> </ul> |
| <ul style="list-style-type: none"> <li>• Se incluyen artículos referidos a juegos cuya interacción está basada en <i>Brain-Computer Interface</i>.</li> </ul>                                                                                             |
| Criterios de exclusión                                                                                                                                                                                                                                    |
| <ul style="list-style-type: none"> <li>• Las soluciones a problemas basadas en el uso de juegos o la investigación en torno a motores de juego no se incluyen cuando no están referidas a la usabilidad del juego en cuestión.</li> </ul>                 |
| <ul style="list-style-type: none"> <li>• La evaluación del rendimiento y la usabilidad de un juego no se incluye cuando está basada en medidas fisiológicas.</li> </ul>                                                                                   |
| <ul style="list-style-type: none"> <li>• No se incluyen los artículos que realizan nuevas propuestas de métodos de evaluación de la usabilidad incluso cuando incluyen una evaluación de usabilidad de ejemplo.</li> </ul>                                |

Tabla 1\_ Criterios de inclusión y exclusión para la búsqueda sistemática

Inicialmente se realiza un análisis descriptivo de los artículos. Los juegos analizados están en su mayoría relacionados con la educación y la salud. Hay otro conjunto notable de juegos que no están en las dos categorías anteriores y que se centran especialmente en la accesibilidad y finalmente un resto de juegos que incluye productos comerciales u otros sin foco específico en ningún área. Los juegos relacionados con la salud se clasifican a su vez en los dedicados a educación sanitaria, formación de pacientes, formación de personal sanitario, diagnóstico médico y una sección miscelánea. Se observa que la mayoría de juegos se orientan a la formación de pacientes y en difusión de educación sanitaria.

La falta de acuerdo en la terminología causa que al extraer de los artículos analizados los términos que describen los aspectos de la usabilidad se quieren evaluar se obtenga una larga lista de términos, en muchos casos semánticamente solapados. En cualquier caso los términos más frecuentes son: satisfacción, acep-

tación, adherencia, jugabilidad, experiencia de usuario, motivación,... lo que da idea de la importancia que tiene el aspecto de la usabilidad que hemos denominado satisfacción frente a los otros dos: eficacia y eficiencia.

También se observa en el análisis una tendencia creciente en la publicación de estudios en torno a la evaluación de la usabilidad de juegos, lo que parece indicar un creciente interés en relación con este problema.

En cuanto al tipo de interfaz que presentan los juegos analizados, un 42% de los estudios se centran en diseños innovadores, como interfaces basadas en la respiración o en la inclusión del espacio físico donde se juega. El conjunto de propuestas se distribuye tal como se muestra en la Figura 3.

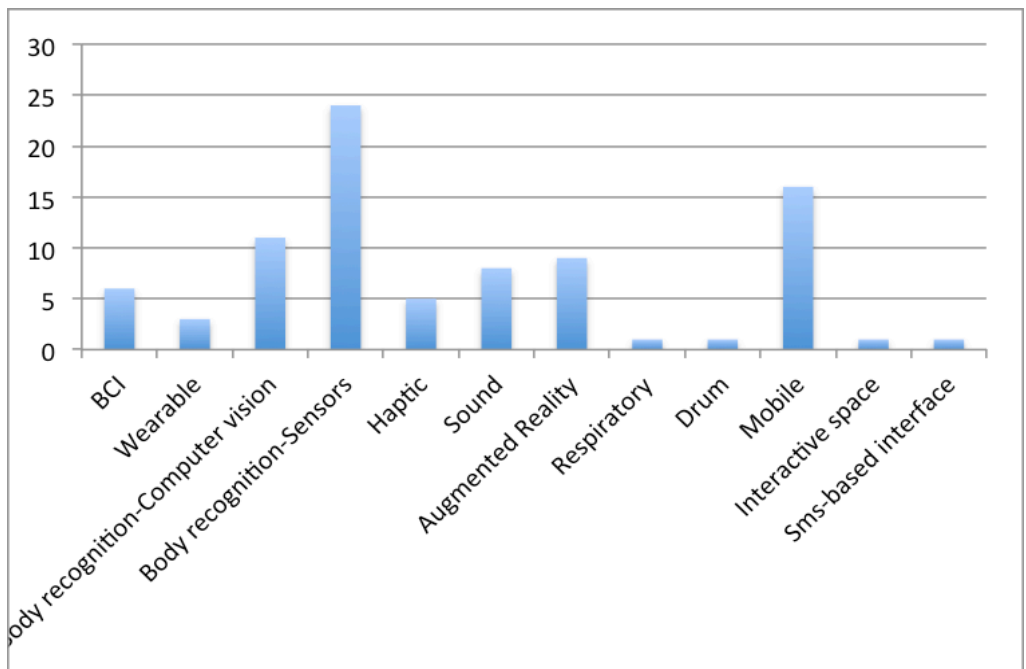


Figura 3\_Interfaces innovadoras utilizadas en un 42% de los estudios

Analizando los métodos de evaluación de la usabilidad utilizados, sólo un 9% de los estudios usan evaluaciones conducidas por expertos. En más de la mitad de los casos (5% del total) se usan de forma complementaria a evaluaciones realizadas involucrando a usuarios. Un 13% de los estudios incluyen más de una evaluación basada en usuarios sobre el mismo juego.

Manejamos la hipótesis de que el uso marginal de evaluaciones conducidas por expertos puede deberse a que es una técnica más adaptada a fases preliminares del desarrollo y la mayoría de las evaluaciones se realizan sobre prototipos completamente funcionales o productos finales. Es imposible concluir si el uso previo de evaluaciones guiadas por expertos pudiera haber mejorado los resultados finales pero es una buena cuestión para guiar investigaciones futuras.

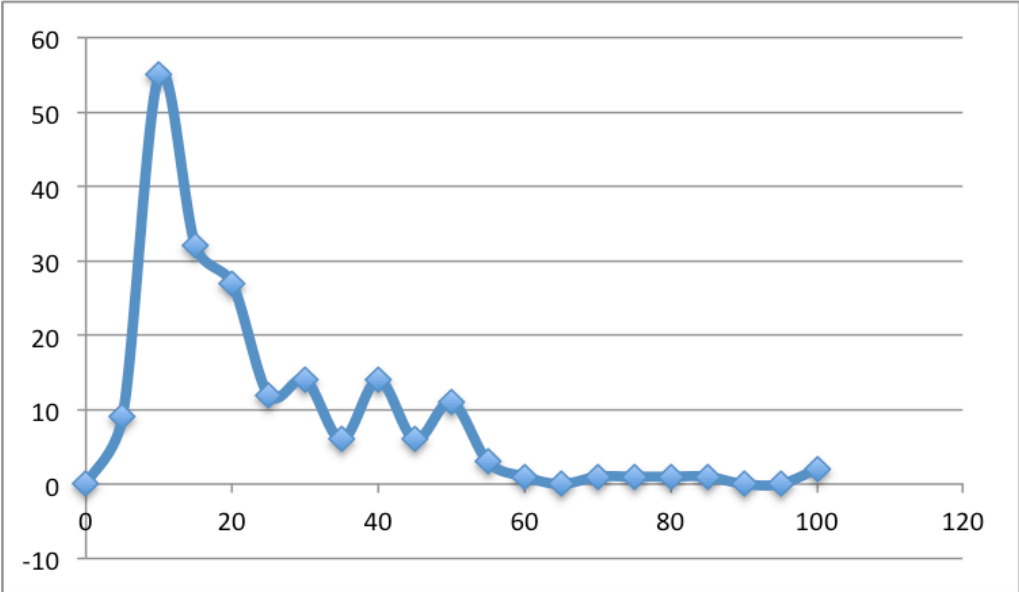


Figura 4\_Histograma de frecuencias del número de usuarios involucrados en las evaluaciones

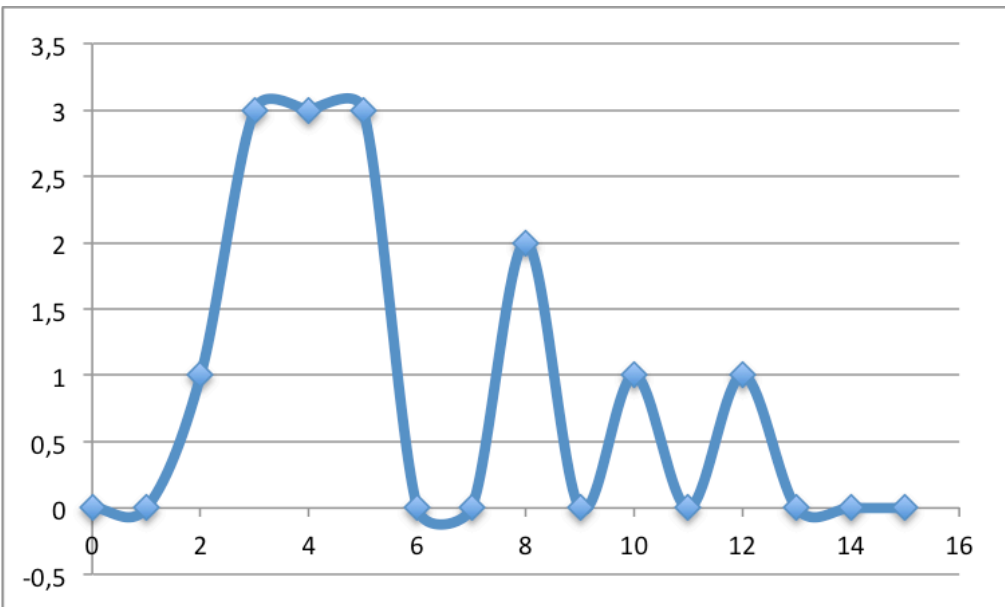


Figura 5\_Histograma de frecuencias del número de expertos involucrados en las evaluaciones

Una caracterización y análisis de qué tipo de usuario se recluta para las evaluaciones ofrece que en la mayor parte de los casos se trata de estudiantes. Resultado lógico tratándose de estudios académicos y siendo el tipo de usuario más accesible a menor coste.

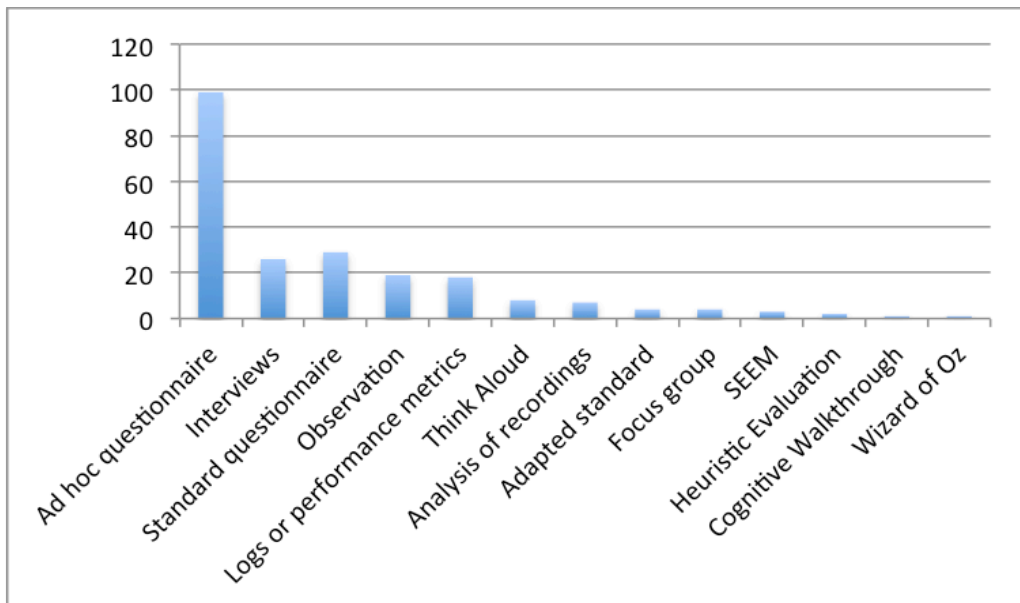


Figura 6\_Técnicas aplicadas en las evaluaciones basadas en *playtesting*

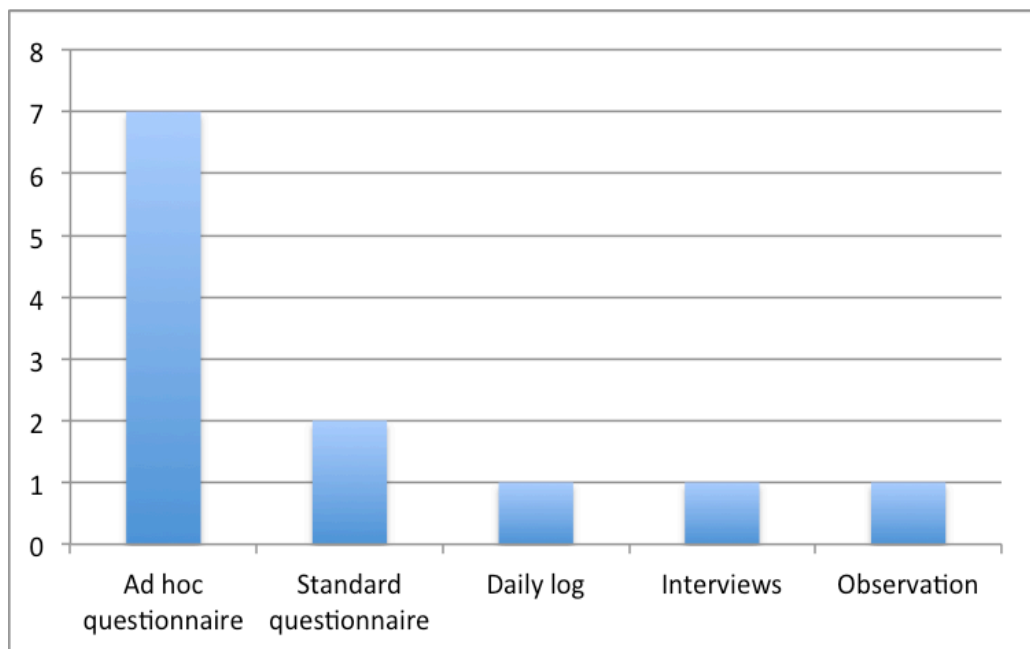


Figura 7\_Técnicas aplicadas en los análisis posteriores al uso

Un aspecto controvertido en la literatura es cuál es el número óptimo de usuarios que debe involucrarse en una evaluación. El criterio de inclusión establecido se corresponde con el acuerdo generalizado sobre el mínimo de usuarios a reclutar, aunque otros autores recomiendan un número mayor de cinco o adaptar el número a la técnica de evaluación usada [15, 48, 49, 50, 51]. El histograma de frecuencias del número de usuarios involucrados en los estudios analizados se muestra en la Figura 4. Como se observa, lo más habitual es incluir a 10, 15 ó 20 usuarios. En el caso de los estudios basados en expertos, un análisis similar determina que lo más frecuente es involucrar a cuatro o cinco expertos (ver Figura 5)

Dentro de los evaluaciones basadas en usuarios se observa que un 93% de ellas recurren al *playtesting*, mientras que un 7% se basan en un análisis posterior al uso de la satisfacción y el rendimiento, siendo este uso más largo en el tiempo al que se realiza en una sesión de *playtesting*. Para ambos casos se muestra la distribución de técnicas utilizadas en las figuras 6 y 7. Estas técnicas se usan tanto individualmente como combinadas.

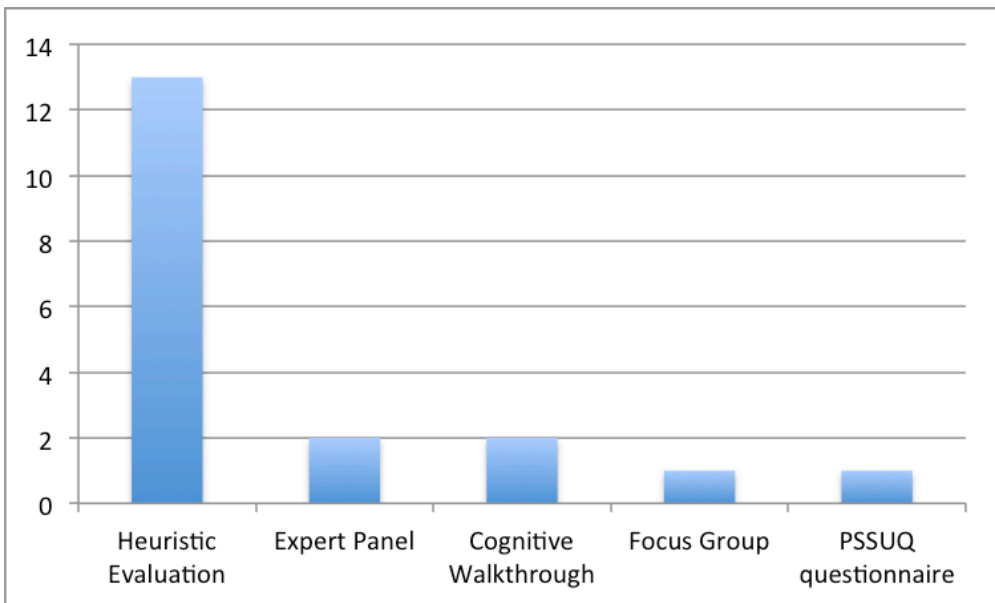


Figura 8\_ Técnicas aplicadas en las evaluaciones basadas en expertos

En el caso de los cuestionarios estándar, se usa con mayor frecuencia, 32% de los casos, la System Usability Scale (SUS) [58] seguida por el End-User Usability Questionnaire [52] en el 5% de los casos y el Nasa Task Load Index Questionnaire (NASA-TLX) [53] en el 7% de los casos.

Un análisis similar en las evaluaciones basadas en expertos muestra que la Evaluación Heurística es el método preferido (ver Figura 8). De entre las herramientas heurísticas disponibles, la propuesta por Nielsen [18] es la más usada. Obsérvese de cara a la interpretación de la Figura 9 que si un método se utiliza varias veces en el mismo artículo sólo se contabiliza una vez.

Tras esta descripción de los hallazgos, se intentan responder las preguntas formuladas como objetivo 5 del análisis. En cuanto a la primera pregunta, la RQ1 (¿Cuál es la aproximación preferida al evaluar la usabilidad de juegos: descartar las técnicas tradicionales y desarrollar nuevos métodos; aplicar directamente herramientas tradicionales estándar bajo la hipótesis de que las diferencias entre juegos y software de productividad no son críticas; o adaptar las herramientas tradicionales existentes para superar dichas diferencias?), se puede concluir que la estrategia más habitual es aplicar técnicas tradicionales en el ámbito de los juegos, no sólo en la evaluación con usuarios sino también en las evaluaciones guiadas por expertos. Esto es así incluso cuando, como dijimos, el 42% de las interfaces son innovadoras y definen nuevos modelos de interacción con los usuarios.

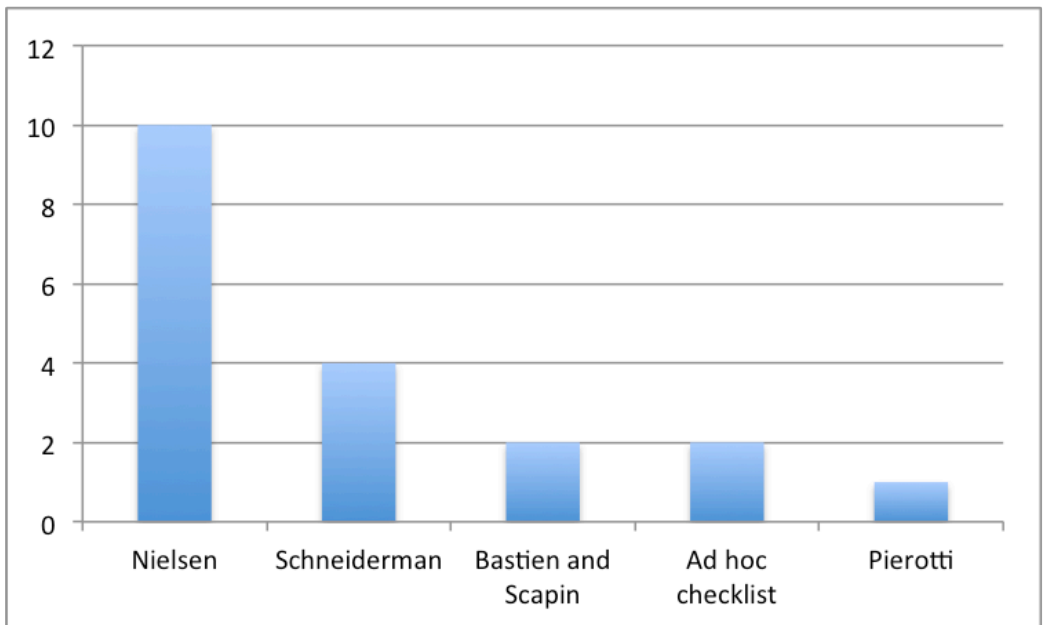


Figura 9\_Listas heurísticas utilizadas en las evaluaciones heurísticas

En cuanto a la cuestión RQ1.1 (¿Se usan técnicas conducidas por tareas para evaluar juegos? ¿Se adaptan de alguna forma?), se observa que en la mayoría de las evaluaciones con usuarios tipo *playtesting* no se establecen tareas a realizar sino que se trata de navegaciones exploratorias por parte de los usuarios. Sin embargo, sí que se recogen métricas de rendimiento en un 10% de las sesiones de

*playtesting*, lo que responde a la cuestión RQ1.2 (¿Se entiende que el concepto de rendimiento en juegos es distinto que en otros dominios? ¿Se redefine el concepto de error?). Los errores son efectivamente redefinidos y se mide el rendimiento en pequeñas tareas para evaluar la claridad de la interfaz y la adecuación de las metáforas pero no se analiza el rendimiento a la hora de realizar tareas complejas o casos de uso. Este tipo de métricas de rendimiento son usadas especialmente cuando el objetivo es comparar dos interfaces alternativas.

En relación a cómo se manejan las interrupciones por parte del instructor que incluyen determinadas técnicas de evaluación y respondiendo por tanto a la cuestión RQ1.3 (¿Las interrupciones del instructor se evitan conscientemente durante el test con usuarios? ¿Se utiliza la técnica *Think Aloud* de forma frecuente?), la técnica *Think Aloud* se usa en un 5% de los casos, combinada o no con otras técnicas. Sin embargo, la conveniencia de usar o no esta técnica está directamente relacionada con el tipo de usuarios involucrados en la evaluación, sus capacidades cognitivas y el entorno en el que se desarrolla el test.

Respecto a la pregunta RQ1.4 (Se mide la satisfacción con una atención especial? ¿Se usan las nuevas propuestas de heurísticas para este objetivo?), el método más prevalente a la hora de medir la satisfacción es el uso de cuestionarios después de sesiones de *playtesting* o uso. Sin embargo, los cuestionarios estándar (los únicos validados estadísticamente) sólo se usan en un 17% de los casos frente al 58% de sesiones tras las que se utiliza un cuestionario *ad hoc*. Pareciera que a pesar de su validez estadística, los cuestionarios estándar no se consideran capaces de atender a los aspectos más controvertidos de los juegos.

Y finalmente, la pregunta RQ2 (¿Cuáles son las características más destacables de los juegos evaluados y del entorno de testeo en las evaluaciones que se describen en la literatura?) lleva a hallazgos interesantes, algunos ya destacados en la propia descripción del conjunto de estudios. Destacaremos el hecho de que las áreas más interesadas en la evaluación de usabilidad en juegos están relacionadas con el aprendizaje y la salud. El uso de juegos serios en estas áreas está actualmente siendo explorado y las características especiales de los usuarios finales hace imprescindible una atención específica a la usabilidad. Por otra parte, la satisfacción es un objetivo prioritario en las evaluaciones de usabilidad analizadas, quizá también por su efecto motivador y de adherencia de cara a los usuarios de los juegos serios.

A pesar de que la revisión sistemática llevada a cabo deja abiertas distintas preguntas y áreas de investigación futuras, presenta una visión global útil para futuros trabajos en el área de evaluación y aseguramiento de la usabilidad en juegos.



## Bloque 3: Evaluación Heurística de juegos: un enfoque modular

Para desarrollar la herramienta MUSE, capaz de producir una lista heurística personalizada para un plan de evaluación específico (objetivo 6) fue necesario establecer un marco de análisis y clasificación de las heurísticas ya propuestas en la bibliografía para el entorno específico de juegos que fue el elegido como caso de ejemplo. Para ello se llevaron a cabo los siguientes trabajos:

1. Seleccionar y filtrar un subconjunto de heurísticas desde la bibliografía para usarlas como motor de heurísticas desde las que realizar las selecciones pertinentes.
2. Definir un conjunto de metadatos aplicables al conjunto de heurísticas obtenidos en el paso anterior y que habiliten los procesos de selección de elementos en función de las características del plan de evaluación.
3. Modelado de la herramienta de acuerdo con el mapa de etiquetas de metadatos y la lista de heurísticas compiladas en el motor. La herramienta obtenida en este punto se considera un prototipo ampliable que permite evaluar la viabilidad de los análisis realizados y que se prevé que evolucione con la adición de nuevas listas heurísticas, potencialmente aplicables a otros entornos específicos además de los juegos; nuevos metadatos, o nubes de metadatos si se trata de nuevos entornos de aplicación; y con la creación de una interfaz gráfica que facilite su manejo y difusión.

La selección de heurísticas se realizó con el objetivo de incluir las listas más referenciadas de entre las propuestas para el entorno específico de los juegos. Así el motor de la herramienta MUSE incluye las propuestas de Federoff [34], Korhonen and Koivisto [35], Pinelle, Wong and Statch [31] and Desurvire et al.[23], así como los trabajos propuestos por estos mismos autores orientados a evolucionar sus propuestas iniciales o corregir carencias de las mismas [54, 55, 56]. Adicionalmente, se han incluido dos trabajos con enfoques y objetivos más específicos teniendo en cuenta nuevos géneros o plataformas [23, 57]; de esta forma la selección final abarca un subconjunto más representativo de juegos. En cualquier caso, la selección se entiende como preliminar y queda abierta la adición de nuevas propuestas al motor central de elementos de Evaluación Heurística en las siguientes iteraciones de la herramienta.

De la selección anterior se obtuvieron un total de 237 heurísticas que fue necesario filtrar para evitar elementos duplicados o solapados. También se descompusieron las preguntas múltiples y finalmente, se adaptó la redacción para

hacerla homogénea. Tras este proceso de filtrado el motor inicial de la primera iteración de desarrollo de MUSE contiene 160 reglas heurísticas.

Para definir los metadatos a aplicar sobre las heurísticas anteriores y poder establecer selecciones en función de las características de un plan de evaluación concreto se tuvieron en cuenta tres cuestiones fundamentales a las que debería atender cada plan de evaluación:

1. Qué aspectos de la definición de usabilidad van a ser objeto de evaluación.
2. Qué tipo de juego va a ser evaluado y cuáles son sus características específicas.
3. Cuál es la fase de desarrollo del juego a evaluar, sobre todo en cuanto a la cercanía al producto final del prototipo evaluado.

Pero la definición de la nube de metadatos debe tener en cuenta, además de los aspectos comunes de un plan de evaluación, qué categorías semánticas pueden superponerse a las distintas propuestas de listas de heurísticas. Estas categorías semánticas atienden a distintos criterios de clasificación. Quizá la conclusión más interesante en este proceso es el haber desechado el género de un juego como una categoría de clasificación útil a la hora de elegir herramientas de análisis de su usabilidad y cómo parece arrojar mejores resultados el descomponer las características esenciales subyacentes que, en muchos casos, son compartidas o combinadas entre juegos de distinto género haciendo que en este sentido los límites de la categorización por género resulten difusos.

De este modo la nube final de metadatos resultante es la mostrada en la Figura 10 y puede observarse cómo el primer nivel de clasificación atiende a los tres aspectos antes destacados desde los planes de evaluación. Una división posterior en categorías atiende a la semántica de las distintas propuestas de la bibliografía.

Como ya se había destacado, esta nube de etiquetas o metadatos se concibe como una primera aproximación apropiada para una primera iteración de MUSE pero en su diseño se ha tenido en cuenta la capacidad de extensión, futura actualización y facilidad de manejo. Debería ser fácil añadir nuevas etiquetas y nuevas categorías que mantuvieran la coherencia del conjunto y ampliaran las capacidades de clasificación y selección en futuras versiones de la herramienta.

Una vez seleccionadas y filtradas las fuentes de datos y creada la nube de etiquetas que permite su clasificación, se ha modelado una herramienta que mantiene capacidades de extensión y ampliación atendiendo, de nuevo, a que es una primera iteración de un desarrollo a más largo plazo.

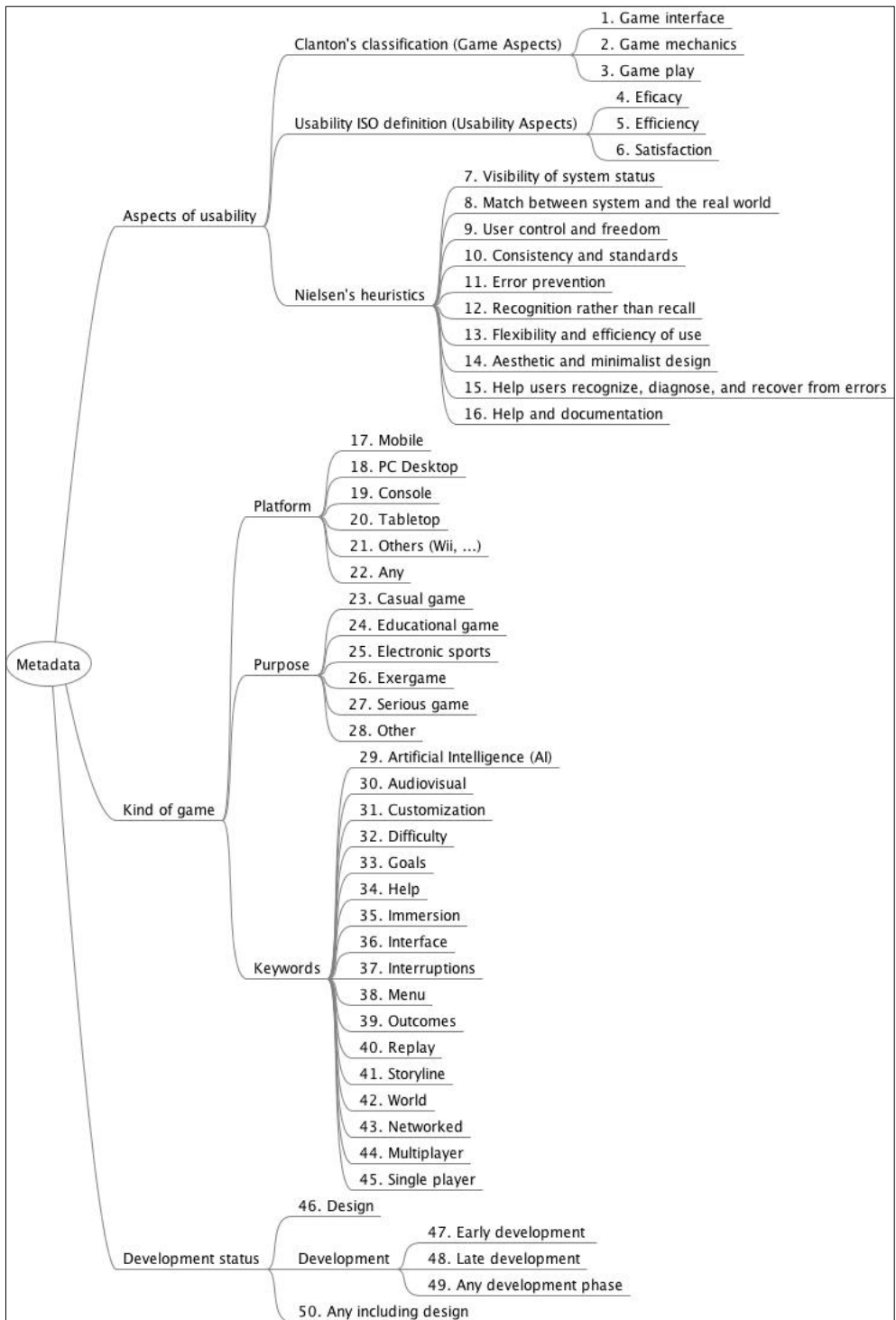


Figura 10\_Nube de metadatos

Volviendo a los objetivos generales, se realizó una evaluación preliminar para valorar por un lado en qué medida el nivel de abstracción de las preguntas presentadas hace la herramienta manejable por evaluadores no expertos (objetivo 7), así como, por otro lado, que la herramienta presenta resultados prometedores en comparación con listas heurísticas generalistas (objetivo 8). Dicha evaluación incluye un piloto inicial con un solo evaluador que permitiera ajustar el análisis y un test posterior con la colaboración de ocho evaluadores siguiendo la recomendación de Hwang and Salvendy [15] y aumentando en dos el número mínimo recomendado por Nielsen [19]. La evaluación contrapone la lista heurística personalizada generada por MUSE para dos juegos de distintas características al uso sobre dichos juegos de la lista generalista propuesta por Nielsen y que, como ya hemos dicho, es la más utilizada incluso en el entorno específico de juegos [17].

Además de medidas cuantitativas relativas al número de errores de usabilidad detectados y su descripción estadística, se establecieron dos medidas adicionales: por una parte, y de cara a evaluar la facilidad de uso de la herramienta por parte de los evaluadores no expertos, se usó un test adaptado desde el estándar SUS [58]. De la aplicación de este test parece extraerse la conclusión de que los evaluadores consideran que el manejo de la heurística generalista de Nielsen y el de las producidas por MUSE es similar en términos de efectividad, eficiencia y satisfacción. Esto lleva a la conclusión de que la usabilidad de la propia herramienta generada por MUSE debe mejorarse en las siguientes versiones a fin de cubrir el objetivo 7.

Por otra, se estableció una medida para el «sesgo del evaluador» [59]. Este sesgo se relaciona con el hecho de que varios evaluadores al analizar la misma interfaz con la misma herramienta de evaluación detectan distintos errores de usabilidad. Este sesgo debe ser minimizado para aumentar la confiabilidad de los resultados de la evaluación en el sentido de que distintas evaluaciones deben conducir a resultados similares y en cuanto a que los errores detectados durante la evaluación deben ser los mismos que surgirían en un uso normal de la interfaz. El sesgo del evaluador fue medido usando la métrica *any-two agreement* formulada por Hertzum and Jacobsen [59]. Los resultados fueron positivos y se puede considerar que en la evaluación el sesgo del evaluador se mantuvo en niveles mínimos.

A pesar de que, como decíamos, los evaluadores encontraron que el manejo de la lista heurística de Nielsen y las listas adaptadas ofrecidas por MUSE eran similares en cuanto a usabilidad de la propia herramienta, en los resultados cuantitativos relativos a la detección de defectos de usabilidad se observó que el número de errores detectados usando la lista generalista era menor que cuando se utilizó una lista adaptada. Este resultado que, en principio, podría apuntar a que la herramienta se orienta al cumplimiento del objetivo 8, debe ser matizado debido al sesgo introducido por la falta de experiencia en usabilidad de los evaluadores y

se concluye que en futuras evaluaciones este tipo de evaluador debería recibir una formación previa al uso de la lista de Nielsen cuyos elementos, expresados en términos mucho más abstractos que los de las listas producidas por MUSE, son difíciles de interpretar cuando se carece de conocimientos previos.

También se realizó un análisis cuantitativo destinado a valorar si se habían minimizado el número de falsos positivos y negativos en el uso de la herramienta (objetivo 9), es decir, que se minimizara el número de elementos incluidos en la lista adaptada final que no resultaran no aplicables y también se minimizara la ausencia de elementos que hubieran sido útiles y que no hubieran aparecido en la lista adaptada debido a fallos en el proceso de selección de elementos de la lista heurística. Por diseño de esta primera iteración de la herramienta se decidió minimizar la posibilidad de que no se seleccionara un elemento potencialmente útil a costa de maximizar el número de falsos positivos, es decir, de elementos incluidos en la lista adaptada y finalmente no aplicables. Esto dio lugar a listas muy largas que supusieron el principal obstáculo de usabilidad detectado por los evaluadores a la hora de aplicar las listas generadas por MUSE. Al igual que los resultados anteriores, estas conclusiones deberá guiar la futura evolución de la herramienta.



## Capítulo 4

### Conclusiones

*Make sure everything you do is so completely crazy it's unbelievable.*

*Roald Dahl*

La presente Tesis Doctoral está formada por un compendio de tres artículos que se detallan a continuación junto con las principales conclusiones obtenidas en cada uno de ellos.

#### Artículo 1:

Yáñez Gómez, R., Cascado Caballero, D., & Sevillano, J. L. (2014). Heuristic evaluation on mobile interfaces: A new checklist. *The Scientific World Journal*, 2014. <http://dx.doi.org/10.1155/2014/434326>

Indicios de Calidad: Este artículo, que ha recibido hasta el momento 77 citas (según Google Scholar, Marzo 2019) y ha sido descargado más de 33161 veces (según la web de la editorial), forma parte de un número especial (*Special Issue*) titulado *Intelligent User Interface for Interactive Multimedia: Emerging Techniques and Services*. La revista, *The Scientific World Journal*, tiene en el año de publicación (2014) un índice de impacto SJR=0,35 (Q2). En el momento en que el artículo fue enviado para su posible publicación, la revista ocupaba también el segundo cuartil del índice JCR (*Journal Citation Reports*), con un índice de impacto IF=1,219. Sin embargo, lamentablemente la revista abandonó el índice JCR justo el año en que nuestro artículo fue publicado.

#### Conclusiones:

- Se ha construido y analizado una heurística específica para el contexto móvil que constituye una herramienta de evaluación de alto detalle que permite evaluar el diseño inicial de una interfaz móvil.
- Se ha mostrado que es posible reutilizar heurísticas en nuevos contextos tecnológicos en base al hecho de que las heurísticas derivan del comportamiento humano y no de la tecnología<sup>[60]</sup>
- La lista heurística generalista propuesta por Nielsen se ha completado con propuestas específicas para un nuevo contexto con el objetivo de cubrir aspectos concretos de la interacción.

- Todo apunta a que puede ser útil ponderar las heurísticas en función del aspecto de la interacción al que aplican.

## Artículo 2:

Yanez-Gomez, R., Cascado-Caballero, D., & Sevillano, J. L. (2017). Academic methods for usability evaluation of serious games: a systematic review. *Multimedia Tools and Applications*, 76(4), 5755-5784.. <https://doi.org/10.1007/s11042-016-3845-9>

Indicios de Calidad: Este artículo formó parte del *Special Issue* titulado *Interactive Media: Technology and Experience* de la revista *Multimedia Tools and Applications* (MTAP) de la editorial Springer. Ha recibido hasta el momento 13 citas (según Google Scholar, marzo 2019), y ha sido descargado más de 1100 veces (según la web de la editorial) a pesar de no estar disponible como open-access. La revista ocupa la posición 42 de 104 (segundo cuartil) de los JCR en la categoría *Computer Science – Software Engineering*, con índice de impacto IF=1,541 (2017).

## Conclusiones:

- La revisión sistemática llevada a cabo muestra el estado de las evaluaciones de usabilidad en el ámbito académico y las principales tendencias a la hora de elegir herramientas para llevarlas a cabo.
- El interés en evaluar la usabilidad de los juegos es una tendencia creciente a pesar de la falta de consenso en la terminología del corpus de técnicas a aplicar, y esto es especialmente significativo en el caso de juegos serios aplicados a las áreas de la salud y la docencia.
- Los aspectos de usabilidad a los que se atiende con más interés en las evaluaciones son los motivacionales, es decir, aquellos relacionados con la satisfacción; entendida ésta como el tercer aspecto, junto con eficacia y eficiencia, de la definición de usabilidad. Sin embargo, los aspectos motivacionales han sido tradicionalmente desplazados por las herramientas genéricas frente a los cognitivos, más estrechamente ligados con el proceso de interacción hombre-máquina.
- Los métodos basados en expertos se usan minoritariamente (9% de los casos) a la hora de evaluar la usabilidad de los juegos, a pesar de sus potenciales ventajas.
- De entre las herramientas que involucran usuarios, los cuestionarios confeccionados ad-hoc constituyen la técnica más prevalente a la hora de evaluar la usabilidad de los juegos. Y esto a pesar de ser herramientas no es-



tandarizadas de las que se extrae información que no es comparable ni válida estadísticamente.

- De entre las metodologías basadas en el criterio de expertos, la Evaluación Heurística es, con diferencia, la más usada a la hora de evaluar los juegos. Aunque su
- Y de entre todas las propuestas de listas heurísticas de la literatura, la propuesta por Nielsen[18] es, con diferencia, la más usada incluso tratándose de la aplicación de una heurística genérica a un contexto específico para el que existen en la literatura numerosas propuestas adaptadas.

### Artículo 3:

Yanez-Gomez, R., Font, J. L., Cascado-Caballero, D., & Sevillano, J. L. (2018). Heuristic usability evaluation on games: a modular approach. *Multimedia Tools and Applications*, 1-28. <https://doi.org/10.1007/s11042-018-6593-1>

Indicios de Calidad: Este artículo forma parte del *Special Issue* titulado *Next Applications in Multimedia Computing* de la revista *Multimedia Tools and Applications (MTAP)*. Ha sido descargado 157 veces (según la web de la editorial) a pesar de no estar disponible como open-access y de que ha sido asignado a un número definitivo muy recientemente (febrero de 2019). Los datos más recientes disponibles de MTAP son los indicados anteriormente, es decir, ocupa la posición 42 de 104 (segundo cuartil) de los JCR en la categoría *Computer Science – Software Engineering*, con IF=1,541 (2017).

### Conclusiones:

- Se ha desarrollado una herramienta denominada MUSE (*Meta-beUristics uSability Evaluation tool*) capaz de producir una lista heurística personalizada para un plan de evaluación específico. Para ello, se seleccionó y filtró un subconjunto de heurísticas desde la bibliografía, y se definió un conjunto de metadatos aplicables a este conjunto de heurísticas que permite seleccionar elementos en función de las características del plan de evaluación.
- Una evaluación preliminar de la herramienta MUSE demuestra que ofrece resultados prometedores a la hora de mejorar la detección de vulneraciones de la usabilidad frente al uso de herramientas genéricas en el contexto de los juegos y cuando los evaluadores no son expertos.
- Adicionalmente, el uso de la herramienta parece reducir el “efecto del evaluador” aumentando la confiabilidad de los resultados.



## Capítulo 5

### Bibliografía

*If I have seen further  
it is by standing on the shoulders of Giants.*  
Newton

- [1] Chung, L., Nixon, B. A., Yu, E., & Mylopoulos, J. (2012). *Non-functional requirements in software engineering* (Vol. 5). Springer Science & Business Media.
- [2] Stallings, W., Brown, L., Bauer, M. D., & Bhattacharjee, A. K. (2012). *Computer security: principles and practice* (pp. 978-0). Upper Saddle River (NJ): Pearson Education.
- [3] ISO, S. 9241-11 (1998) Ergonomic requirements for office work with visual display terminals (VDTs)-Part II guidance on usability
- [4] i Saltiveri, T. G., Vidal, J. L., & Delgado, J. J. C. (2011). *Diseño de sistemas interactivos centrados en el usuario*. Editorial UOC.
- [5] ISO, S. 71 (2014) Guide for addressing accessibility in standards.
- [6] Karat, J. (1997). User-centered software evaluation methodologies. In *Handbook of human-computer interaction* (pp. 689-704). North-Holland.
- [7] Dongsong Zhang. *Overview of usability evaluation methods*. <http://www.usabilityhome.com> Retrieved in March 2014
- [8] Heo, J., Ham, D. H., Park, S., Song, C., & Yoon, W. C. (2009). A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors. *Interacting with Computers*, 21(4), 263-275.
- [9] Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys (CSUR)*, 33(4), 470-516.
- [10] Po, S., Howard, S., Vetere, F., & Skov, M. B. (2004, September). Heuristic evaluation and mobile usability: Bridging the realism gap. In *International Conference on Mobile Human-Computer Interaction* (pp. 49-60). Springer, Berlin, Heidelberg.
- [11] Bertini, E., Gabrielli, S., Kimani, S., Catarci, T., & Santucci, G. (2006, May). Appropriating and assessing heuristics for mobile computing. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 119-126). ACM.
- [12] Duh, H. B. L., Tan, G. C., & Chen, V. H. H. (2006, September). Usability evaluation for mobile device: a comparison of laboratory and field tests. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 181-186). ACM.
- [13] Nielsen, J. (1994, April). Usability inspection methods. In *Conference companion on Human factors in computing systems* (pp. 413-414). ACM.

- [14]Nielsen, J., & Landauer, T. K. (1993, May). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 206-213). ACM.
- [15]Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: the 10±2 rule. *Communications of the ACM*, 53(5), 130-133.
- [16]Nielsen, J. (1994). Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. *Cost-justifying usability*, 245-272.
- [17]Yanez-Gomez, R., Cascado-Caballero, D., & Sevillano, J. L. (2017). Academic methods for usability evaluation of serious games: a systematic review. *Multimedia Tools and Applications*, 76(4), 5755-5784.
- [18]Nielsen, J., & Molich, R. (1990, March). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249-256). ACM.
- [19]Nielsen, J. (1994). *Usability engineering*. Elsevier.
- [20]Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., & Ames, M. (2003, April). Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 169-176). ACM.
- [21]November 2013, <http://innovation.logica.com.es/web/procura>.
- [22]Yáñez Gómez, R., Cascado Caballero, D., & Sevillano, J. L. (2014). Heuristic evaluation on mobile interfaces: A new checklist. *The Scientific World Journal*, 2014.
- [23]Desurvire, H., Caplan, M., & Toth, J. A. (2004, April). Using heuristics to evaluate the playability of games. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1509-1512). ACM.
- [24]Malone, T. W. (1982, March). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on Human factors in computing systems* (pp. 63-68). ACM.
- [25]Torrente, M. D. C. S. (2011). *SIRIUS: Sistema de Evaluación de la Usabilidad Web Orientado al Usuario y basado en la Determinación de Tareas Críticas* (Doctoral dissertation, Universidad de Oviedo).
- [26]Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview.
- [27]Ang, C. S., Avni, E., & Zaphiris, P. (2008). Linking pedagogical theory of computer games to their usability. *International Journal on E-Learning*, 7(3), 533-558.
- [28]Lewis, C., Polson, P. G., Wharton, C., & Rieman, J. (1990, March). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 235-242). ACM.
- [29]Bias, R. (1991). Interface-Walkthroughs: efficient collaborative testing. *IEEE Software*, 8(5), 94-95.
- [30]Annett, J., & Duncan, K. D. (1967). Task analysis and training design.

- [31]Pinelle, D., Wong, N., & Stach, T. (2008, April). Heuristic evaluation for games: usability principles for video game design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1453-1462). ACM.
- [32]Moreno-Ger, P., Torrente, J., Hsieh, Y. G., & Lester, W. T. (2012). Usability testing for serious games: Making informed design decisions with user data. *Advances in Human-Computer Interaction, 2012*, 4.
- [33]Clanton, C. (1998, April). An interpreted demonstration of computer game design. In *CHI 98 conference summary on Human factors in computing systems* (pp. 1-2). ACM.
- [34]Federoff, M. A. (2002). *Heuristics and usability guidelines for the creation and evaluation of fun in video games* (Doctoral dissertation, Indiana University).
- [35]Korhonen, H., & Koivisto, E. M. (2006, September). Playability heuristics for mobile games. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 9-16). ACM.
- [36]Caillois, R. (2001). *Man, play, and games*. University of Illinois Press.
- [37]Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive science*, 5(4), 333-369.
- [38]Fabricatore, C., Nussbaum, M., & Rosas, R. (2002). Playability in action videogames: A qualitative design model. *Human-Computer Interaction, 17*(4), 311-368.
- [39]FDIs, I. S. O. (2009). 9241-210: 2009. Ergonomics of human system interaction-Part 210: Human-centered design for interactive systems (formerly known as 13407). International Organization for Standardization (ISO). Switzerland.
- [40]Johnson, D., & Wiles, J. (2003). Effective affective user interface design in games. *Ergonomics, 46*(13-14), 1332-1345.
- [41]Cheng, K., & Cairns, P. A. (2005, April). Behaviour, realism and immersion in games. In *CHI'05 extended abstracts on Human factors in computing systems* (pp. 1272-1275). ACM.
- [42]Koeffel, C., Hochleitner, W., Leitner, J., Haller, M., Geven, A., & Tscheligi, M. (2010). Using heuristics to evaluate the overall user experience of video games and advanced interaction games. In *Evaluating user experience in games* (pp. 233-256). Springer, London.
- [43]Yanez-Gomez, R., Font, J. L., Cascado-Caballero, D., & Sevillano, J. L. (2018). Heuristic usability evaluation on games: a modular approach. *Multimedia Tools and Applications*, 1-28.
- [44]Myers, D. (2003). *The Nature of Computer Games: Play As Semiosis (Digital Formations, V. 16)*. Peter Lang Publishing.
- [45]Korhonen, H. (2011, November). The explanatory power of playability heuristics. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (p. 40). ACM.
- [46]Johnson, D., & Wiles, J. (2003). Effective affective user interface design in games. *Ergonomics, 46*(13-14), 1332-1345.

- [47] Desurvire, H., & Wiberg, C. (2009, July). Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration. In *International conference on online communities and social computing* (pp. 557-566). Springer, Berlin, Heidelberg.
- [48] Jiménez, C., Rusu, C., Roncagliolo, S., Inostroza, R., & Rusu, V. (2012, November). Evaluating a methodology to establish usability heuristics. In *2012 31st International Conference of the Chilean Computer Science Society* (pp. 51-59). IEEE.
- [49] Tullis, T. S., & Stetson, J. N. (2004, June). A comparison of questionnaires for assessing website usability. In *Usability professional association conference* (Vol. 1).
- [50] Nielsen, J. (2012). Why you only need to test with 5 users, 2000. <https://www.nngroup.com>. Retrieved in March 2019
- [51] Preece J., Sharp H., Rogers Y. (2015) Interaction design-beyond human-computer interaction. Wiley
- [52] Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough?. *Human factors*, 34(4), 457-468.
- [53] Sánchez, J. (2003). End-user and facilitator questionnaire for Software Usability. *Usability evaluation test. University of Chile, Santiago, Chile.*
- [54] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.
- [55] Korhonen, H., & Koivisto, E. M. (2007, September). Playability heuristics for mobile multi-player games. In *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts* (pp. 28-35). ACM.
- [56] Pinelle, D., Wong, N., Stach, T., & Gutwin, C. (2009, May). Usability heuristics for networked multiplayer games. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 169-178). ACM.
- [57] Desurvire, H., & Wiberg, C. (2009, July). Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration. In *International conference on online communities and social computing* (pp. 557-566). Springer, Berlin, Heidelberg.
- [58] Papaloukas, S., Patriarcheas, K., & Xenos, M. (2009, September). Usability assessment heuristics in new genre videogames. In *2009 13th Panhellenic Conference on Informatics* (pp. 202-206). IEEE.
- [59] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- [60] Hertzum, M., & Jacobsen, N. E. (2001). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, 13(4), 421-443.
- [61] Budiu, R., & Nielsen, J. (2011). Usability of ipad apps and websites.

## **Parte II**

### **Publicaciones**





Artículo 1

# Heuristic Evaluation on Mobile Interfaces: A New Checklist

Yáñez Gómez, R., Cascado Caballero, D., & Sevillano, J. L. (2014). Heuristic evaluation on mobile interfaces: A new checklist. *The Scientific World Journal*, 2014.



## Artículo 2

# Academic methods for usability evaluation of serious games: a systematic review

Yanez-Gomez, R., Cascado-Caballero, D., & Sevillano, J. L. (2017). Academic methods for usability evaluation of serious games: a systematic review. *Multimedia Tools and Applications*, 76(4), 5755-5784.



### Artículo 3

## Heuristic evaluation on games: a modular approach

Yanez-Gomez, R., Font, J. L., Cascado-Caballero, D., & Sevillano, J. L. (2018). Heuristic usability evaluation on games: a modular approach. *Multimedia Tools and Applications*, 1-28.

