

Evaluación de aplicaciones para un patrimonio inclusivo: el proyecto EDU2014-52675-R¹.

Mikel Asensio, Joan Santacana & Elena Asenjo.

1-. La falta de cultura de evaluación y la práctica de la monitorización.

En el ámbito de la cultura la evaluación es una *rara avis*, y la evaluación de aplicaciones en el ámbito del patrimonio no es una excepción. Es cierto que hablar de patrimonio supone referirse a un conjunto muy amplio de instituciones en el mundo de la cultura y que van cambiando con el tiempo y con las mentalidades de gestión (Santacana & Hernández, 2006; Santacana & Llonch, 2016; Asensio & Pol, 2017), pero no es menos cierto que todas ellas comparten el ser espacios de contacto con la cultura material e inmaterial a través de un formato expositivo. La mayoría de las implementaciones de los dispositivos no incluyen un plan de evaluación. Lo que es más habitual en estos ámbitos profesionales es un tipo de evaluación formativa que suelen denominar “monitorización” o “monitoreo” (*monitoring*), que suele describirse como pruebas sucesivas que controlan que la aplicación “corre correctamente”, es decir, que pasa de una pantalla a otra cuando se realiza determinada acción. Lo que se comprueba es que la aplicación se va desplegando tal como está previsto en la mente del diseñador. La monitorización la desarrollan, la mayoría de las veces, el mismo equipo de técnicos (generalmente informáticos) que se encargan de su diseño. Y se realiza normalmente en el lugar del diseño de la aplicación (y no in situ, donde habitualmente cambian notablemente las condiciones técnicas). Además, por lo que nosotros mismos hemos podido comprobar en el diseño de este tipo de aplicaciones (cada vez más abundantes en todos los ámbitos, Sada, 2012), un dato a tener en cuenta es que la mayor parte del proceso de monitorización se realiza en pantalla de ordenador y no en un dispositivo final similar a los de los usuarios. Es cierto que suele haber una monitorización final, más o menos in situ y más o menos con dispositivos similares, pero esta monitorización final se vuelve a limitar a que la aplicación corra adecuadamente, es decir, que una persona que la conoce perfectamente y que conoce muy bien qué hay que hacer con ella en cada caso (en cada pantalla) vaya saltando de una acción a la siguiente según el plan previsto por el diseño de la aplicación.

El proceso de monitorización no introduce la variable de usuario. Ni siquiera introduce vías de utilización que incluyan los “usos perversos” que la aplicación permite, es decir, ante una disyunción entre varias acciones el experto siempre optará por las opciones correctas o no correctas previstas, pero el usuario podrá optar entre todas las opciones posibles, aunque sean totalmente disruptivas con la tarea, con sus objetivos o con los intereses en cada caso. Si una acción es posible, los usuarios pueden

¹ Proyecto de I+D+I, del Programa Estatal de Investigación, Desarrollo e Innovación orientada a los Retos de la Sociedad, financiado por la Dirección General de Investigación Científica y Técnica del MINECO. Titulado: ANALISIS EVALUATIVO DE APLICACIONES PARA M-LEARNING DE CARACTER INCLUSIVO EN ESPACIOS PATRIMONIALES. Proyecto: EDU2014-52675-R. Coordinado por Joan Santacana.

realizarla. Además, en muchos casos las *affordances* del sistema llevan al usuario a realizar estas acciones perversas o disruptivas con mucha más frecuencia de la que el experto presupone. Así que una primera diferencia entre el monitoreo y la evaluación es tener en cuenta no solamente la manipulación dirigida del experto, sino cualquier posibilidad de manipulación diversificada que permita la aplicación y el dispositivo. En segundo lugar, el usuario introduce una variable fundamental ya que decidirá utilizar la aplicación en una o en otra dirección y no necesariamente con los motivos previstos coherentemente en el diseño, sino con un conjunto sustancial de otro tipo de posibilidades que no han sido previstas en el diseño porque no se realizan evaluaciones previas de los intereses y disposiciones de los usuarios.

A todo esto, en tercer lugar, hay que sumar que la monitorización no suele tener nunca en cuenta la utilización en contexto, con lo que se pierden la mayoría de las limitaciones de las acciones en los espacios reales. El contexto incluye al menos dos grandes tipos de variables, uno debido a la composición del espacio y los elementos estructurales de la exposición, del museo o del yacimiento. El contexto incluye una gran cantidad de situaciones que condicionan el uso de la aplicación y la relación entre lo que aorta la aplicación y la realidad. Resolver este tipo de problemas es central para que usuario no se vea perdido y para que la aplicación sea fiable.

En segundo lugar, el contexto incluye también la dimensión grupal/social, muy pocas aplicaciones incluyen aún una mínima planificación de las interacciones del usuario con otros usuarios o con actividades de grupo.

La diferencia entre monitorización y evaluación (en cualquiera de sus modalidades) es que la evaluación es un control real del uso de la aplicación con usuarios reales y en el contexto natural.

La evaluación es fundamental para desarrollar la tecnología como una herramienta cultural más del museo inclusivo (Santacana & Coma, 2014; Santacana & López, 2015; Asensio, Santacana y Fontal, 2017; Santacana, J., Martínez, T. & Asensio, 2016).

2-. A años luz de Netflix.

Todo el mundo conoce la plataforma Netflix de televisión a la carta. La compañía tiene un sobresaliente sistema de evaluación que le permite ir adaptando progresivamente los contenidos a las necesidades de los usuarios en su conjunto y, en particular, ir singularizando la oferta al perfil de usuario. Para ello, Netflix tiene un sistema de evaluación bastante complejo que, a juzgar por los resultados obtenidos, diríamos que resulta bastante exitoso. Habitualmente se dice que es un sistema de Big Data, porque muchos identifican los Big Data con aquellos estudios en que se manejan grandes cantidades de datos. Por ejemplo, algunos consideran Big Data a los estudios que llegan a manejar cantidades de datos medidas en 'Petabites' (10^{15} bites). Pero esto es muy simplista. Es obvio que los Big Data utilizan cantidades ingentes de datos. Netflix superó los 100 millones de usuarios el julio pasado, por tanto, si almacenamos un número considerable de datos de cada usuario, y lo cruzamos con un número nada desdeñable de datos de nuestra oferta, la progresión resulta espectacular. Pero no es menos cierto que algunos estudios estadísticos clásicos, por ejemplo, los estudios epidemiológicos, también utilizan enormes cantidades de datos y no por eso son una

metodología diferenciada de los que utilizan una muestra o un volumen de datos menor. Lo que caracteriza al Big Data como sistema de evaluación es que las decisiones se toman en base a análisis complejos que se basan precisamente en que disponemos de una gran cantidad de información y no en meras acumulaciones de datos. Las muestra mega-grandes precisan de procesos de toma de decisión propios que no son los mismos criterios que aplicamos a las muestras grandes.

Por una parte, Netflix analiza el comportamiento real de cada usuario, con medidas sobre lo que ve cada usuario: horarios, temáticas, tipologías de programas, si se ven enteros o se dejan a medias, si ésta actitud se sostiene en el tiempo, si se asocian las elecciones a terceras variables, y un enorme etcétera. Netflix realiza preguntas concretas como por ejemplo rastrear las elecciones de películas infantiles en horario escolar y en horario no escolar, para ver si se eligen precisamente estas temáticas con o sin niños en la casa. O preguntas mucho más sofisticadas como por ejemplo cuál es el número medio de capítulos que un usuario tiene que ver de una serie para engancharse y ver todos los capítulos restantes. Netflix ha sido capaz de esta manera de detectar conductas insólitas y aparentemente disruptivas, por ejemplo, una que nos llama especialmente la atención es que cerca de un 7% de los usuarios ven ciertos programas ¡a doble velocidad! y paran solamente para ver ciertos trozos (por ejemplo si se han engancho con una de las varias historias que tiene la serie o bien solamente con uno de los personajes). Y han detectado también que este comportamiento va en aumento.

Por otro lado, Netflix analiza cuidadosamente las emisiones. Dispone de un “ejército de etiquetadores entrenados para calificar los contenidos en un conjunto cada vez más amplio de categorías que van más allá de los géneros o las tipologías más habituales, por ejemplo, película con final feliz, con actores ingleses, de protagonista joven, con director asiático, con imágenes de países exóticos, sin escenas de terror, ... y un considerable etcétera. El análisis del acuerdo inter-jueces garantiza el ajuste de las categorizaciones. Este etiquetado permite un análisis exhaustivo del ‘imput’. Esta finura es impensable en un mundo como el del patrimonio, en el que se emplean categorizaciones muy simples y superficiales, cuando no ambiguas, tanto sobre los contenidos como sobre los montajes museográficos, lo cual es extensivo a las características de las APPs.

El cruce masivo de ambas fuentes de variables es lo que permite analizar las tendencias. Pero, además, el cruce particular de ambos tipos de variables permite singularizar para cada usuario una oferta determinada.

Es obvio que las APPs de patrimonio no se mueven en estos volúmenes, por tanto no estamos poniendo el énfasis en la utilización de los Big Data como método. Pero sí queríamos llamar la atención sobre la utilidad de la finura de análisis por un lado de los datos reales de usuario y por otro del propio imput que despliega la oferta tecnológica.

3-. Indicadores para expertos y cuestionarios para usuarios.

La escasa evaluación que se realiza suele asentarse en dos modelos diferentes de evaluación, ambas sumativas. Una suele ser la de evaluación cualitativa de indicadores que tradicionalmente se realiza solamente con una pequeña muestra de personas

expertas. La fundamentación de esta práctica hay que buscarla, a nuestro entender, en la tradición de evaluación mediante rúbricas que proviene de los ámbitos educativos (Lee & Cherner, 2015). Saber cómo evaluar un producto de software es tan importante como saber cómo usarlo (Winslow, Dickerson y Lee, 2013). Cada vez son mas las aplicaciones para tabletas (Murray & Olcese, 2011; Pilgrim et al., 2012), y enorme la cantidad de aplicaciones educativas desde hace ya años (Earl, 2013; Rao, 2012). En un estudio seminal, Reeves y Harmon (1993) presentaron 14 dimensiones de evaluación pedagógica y 10 dimensiones de evaluación de interfaz de usuario, y estas dimensiones han servido como base para muchas rúbricas desarrolladas en las últimas dos décadas (por ejemplo, Coughlan & Morar, 2008; Schibeci, et al., 2008).

Pero también en la evolución de las prácticas de monitorización de la propia ingeniería, sin duda presionada por las teorías de diseño, cada vez más exigentes con los aspectos ergonómicos del diseño inteligente. La evaluación mediante guiones de indicadores por personas expertas es una evaluación de carácter racional, basada en el desarrollo de un modelo, y que tiene dos fuentes de justificación, la validez de constructo del modelo desde el punto de vista teórico y práctico, y la fiabilidad inter-jueces (algo en lo que no siempre se pone la suficiente atención). Además, los modelos deberían someterse a una validación empírica, en la que se mida la consistencia interna de la tarea, así como la agrupación empírica de los indicadores en factores estadísticos reales. Nosotros no hemos encontrado en este ámbito ninguna evaluación de indicadores que alcance este nivel. Estos modelos suelen ser bastante inestables ya que evolucionan a medida que se desarrolla la reflexión teórica.

La segunda opción de evaluación suelen ser los cuestionarios. En este caso sí suelen ser contestados por usuarios, aunque no solamente, ya que a veces se incluyen expertos, gestores, diseñadores, etc. Como es sabido, los cuestionarios pueden ser una herramienta muy potente si están bien contruidos. Sin embargo, es cierto que estamos acostumbrados, y este ámbito tampoco es una excepción, a cuestionarios muy superficiales y mal contruidos, que no suelen respetar las condiciones mínimas de fiabilidad y validez, que no explotan las posibilidades de este tipo de herramientas y que no suelen ser posteriormente procesados de manera exhaustiva. En muchas ocasiones se quedan en unos planteamientos muy superficiales, que buscan un análisis confirmatorio de unos pocos intereses concretos y que no van más allá, sin insistir en los aspectos problemáticos o conflictivos, buscando más la auto-complacencia del binomio productor-cliente, que una evaluación real que aporte datos útiles de mejora. Es cierto que la situación real en la que se dan estas evaluaciones condiciona mucho su desarrollo. Por ejemplo, recientemente muchas instituciones incluyen en sus pliegos de contratación la inclusión de algún tipo de evaluación de la gestión. Las empresas se ven así obligadas a informar mediante algún sistema mínimamente creíble de los aspectos de funcionamiento del servicio ofrecido, por tanto, realizar breves cuestionarios que justifiquen una evaluación cuyo único objetivo inamovible es reportarle al cliente que su realización ha sido óptima. Reconozcamos que no es éste el mejor panorama para desarrollar una evaluación ambiciosa.

Los cuestionarios suelen incluir tres tipos de información. Los perfiles (para determinar los tipos de usuario y poder cruzarlo con el resto de variables), el impacto (que busca generalmente unos dígitos cuantitativos) y la opinión (que suele buscar normalmente

los aspectos más cualitativos sobre los aspectos positivos y negativos de la usabilidad entendiéndola en su sentido más amplio. El mayor problema es que los aspectos a evaluar varían mucho de unos cuestionarios a otros, con lo que es muy difícil ir generando una herramienta que comparta unos aspectos comunes. Cada trabajo se fija en aspectos muy diferentes, con concepciones muy distintas de lo que es la usabilidad, y generalmente más pegadas y más interesadas en las cuestiones concretas y en las supuestas aportaciones de cada aplicación y su dispositivo asociado, de manera que se pierde la visión general.

Es cierto que progresivamente podemos encontrar cuestionarios mejor elaborados y que incluyen herramientas cualitativas y cuantitativas algo más sofisticadas (por ejemplo, escalas Likert), que las meras preguntas abiertas o semi-abiertas sin más orientadas por los contenidos concretos.

4-. Cuando la evaluación es una investigación.

En términos estrictos una evaluación no es una investigación, lo mismo que no es una investigación una catalogación o un registro. Si entendemos por evaluar la aplicación de una estrategia fija con un plan prefijado y con una herramienta estándar previamente fijada, la evaluación no es una investigación. Téngase en cuenta que el carácter de investigación de una evaluación no depende de los resultados obtenidos, ni de que estos sean cuantitativos o cualitativos (un error habitual de los comentarios a este tipo de temas, sobre todo en los ámbitos educativos

Indudablemente, el reconocimiento de una evaluación como una investigación le otorga un carácter epistemológico de mayor nivel. La investigación responde preguntas y hace avanzar el conocimiento en sentido constructivo y generativo.

Mientras que la mera evaluación supone mantenerse en un plano acumulativo de mayor información, necesario en muchos casos y muy respetable, no estamos ahora analizando su necesidad o en su pertinencia sino en su potencia.

La investigación sobre la evaluación de las APPs utilizadas en el ámbito del patrimonio se enfrenta entonces al reto de construir, primero, un plan de investigación que genere las preguntas correctas (como todos sabemos, la parte más difícil de un proceso de investigación). Este primer nivel tiene que ver sobre todo con el nivel teórico, es decir, los modelos explicativos con los que nos manejamos y con los que nos podríamos manejar. En nuestro caso, esto supone necesariamente colocarse en un plano interdisciplinar en el que aunar esfuerzos y sensibilidades de muy diversas tradiciones de investigación (al menos ingeniería informática, psicología, educación, patrimonio). Esta confluencia no es sencilla porque además los distintos actores no están igualmente representados en términos de poder a lo largo del proceso de conceptualización, diseño, desarrollo, comercialización y gestión sostenible.

‘*Lazos de luz azul*’ fue un proyecto que desarrollamos hace unos años sobre el uso de la tecnología en el ámbito del patrimonio (Asensio & Asenjo, 2011). En varios de los trabajos allí recogidos se defendía la necesidad de desarrollar evaluaciones más allá de meras acumulaciones de datos descriptivos, con énfasis en datos empíricos que respondieran a preguntas teóricas previas, en el marco de investigaciones desarrolladas desde modelos teóricos definidos. En aquél proyecto pudimos revisar

numerosas investigaciones sobre las páginas web de museos o los formatos de audioguías, el uso de los dispositivos móviles (en aquella época aún PDAs), o nuevas propuestas del uso de *clikers*, o de museos virtuales. En todas ellas y en conjunto, la conclusión matriz fue el convencimiento de una aplicación prematura de los diferentes formatos tecnológicos, mucho antes de que puedan aporten soluciones reales y sostenibles a los problemas planteados en el patrimonio, con una distancia excesiva entre la creación de una expectativas excesivas, usando un discurso grandilocuente y circular, sobre realizaciones que posteriormente decepcionaban por la pobreza de la experiencia transmitida, la inmensa mayoría de las cuáles no contemplaban ningún tipo de evaluación. Finalizábamos el capítulo de conclusiones generales con la siguiente reflexión: “es claro que el desarrollo correcto de las aplicaciones tecnológicas puede ser un elemento central de andamiaje del mensaje expositivo dentro del enfoque de la experiencia museística, siempre que seamos capaces de superar el deslumbramiento inicial de las plataformas tecnológicas en sí mismas y se planifique su utilización dentro del conjunto del sistema de la oferta y reflexión museológica sostenible.”

5-. Encuentros en la tercera fase: contacto

Aun aceptando que la evaluación de indicadores por sujetos expertos y los cuestionarios de usuarios pueden dar una información valiosa cuando están bien hechos. Parece evidente que una evaluación profunda no puede hacerse sin valorar la utilización de la aplicación *'in situ'*. Los evaluadores de exposiciones valoramos sobremanera la herramienta de la observación sobre ninguna otra. La observación proporciona una serie de datos duros sobre los tiempos dedicados por los usuarios a cada una de las partes o acciones de la aplicación, así como los recorridos que realiza, tanto espaciales como en la propia aplicación, al tiempo que permite registrar los comportamientos desarrollados en cada caso, pudiendo distinguir entre las conductas acordes con las opciones de la aplicación y aquellas conductas disruptivas que vienen provocadas por diferentes fuentes de variabilidad, sea el propio sujeto, el contexto, el mensaje patrimonial, el contexto espacial o limitaciones de la propia aplicación. La observación aporta una serie de medidas (accesibilidad, atraktividad, atrapabilidad, absolutas y relativas, tiempos útiles y de transición), perfectamente operativizadas en unidades de tiempo de parada, número de paradas, o pases, delante del *display* o delante de la pantalla concreta o de la acción a realizar, y de una probada consistencia estadística en numerosas evaluaciones (Pol & Asensio, 2017). En la evaluación de exposiciones mediante la técnica de observación, disponemos incluso de meta-análisis aplicados a estas medidas, con la creación de constructos de medida más complejos, que permiten comparar situaciones muy diversas con independencia de su contenido y de sus características formales concretas (Castro, Botella & Asensio, 2016; Asensio et al., 2017). Extrapolando estas medidas parece factible la creación de constructos similares para aplicaciones tecnológicas. Pero se precisa de una medida previa suficientemente potente a nivel de método y de medida como la observación. Si bien muchos colegas remarcan las dificultades de la observación *'in situ'* por razones prácticas no es menos cierto que los beneficios superan las dificultades, y que es

posible realizar simulaciones muy fiables con sujetos relativamente cautivos, confederados o seleccionados de diferentes maneras. Aunque no podamos disponer de todos los datos teóricos que podría aportar una observación, simplemente con disponer de una estimación sumaria de tiempos dedicados y comportamientos desarrollados tendríamos de una información muy valiosa para poderla cruzar con la información proveniente de otros métodos y técnicas.

Por último, destacar que, en cuanto a su calidad, los datos de observación no están sujetos a la distorsión de los datos de auto-informe (por ejemplo, sesgos sociales como la deseabilidad social, o bien sesgos de proceso, como el enmascaramiento); al sesgo que suponen los análisis de los *trackings* de la tecnología que son datos ciegos (en cuanto a la situación en que se produce determinada elección); o al sesgo propio que supone la opinión o la posición teórica de los expertos cuando aplican los indicadores. El segundo tipo de técnicas de evaluación que se precisarían para completar el círculo serían las tareas cuasi-experimentales que nos permiten medir la adquisición de conocimiento, sean de tipo procedimental/competencial, conceptual o emocional. Hay infinitas maneras de medir estos contenidos, con enfoques más cualitativos o más cuantitativos.

La idea básica es la siguiente. Si la aplicación sirve para recorrer determinados contenidos relativos al mensaje patrimonial, cabe preguntarse si los usuarios son capaces de adquirir, reconocer, familiarizarse, manejar, ... o simplemente sensibilizarse, valorar o identificarse con dichos contenidos en mayor o menor medida tras la visita al patrimonio utilizando dicha aplicación.

Las tareas cuasi-experimentales sirven para medir diferentes aspectos de uso, reconocimiento o evocación de dichos contenidos intelectuales y emocionales. Como decimos, hay una infinidad de tareas de estos tipos, las más compartidas con ámbitos de evaluación psicológica o educativa (por ejemplo, tareas de categorización, de memoria o de comprensión), y algunas propias de los ámbitos del patrimonio como, por ejemplo, los '*meaning mapping*' de Falk & Dierking (1992) o las tareas de claves expositivas (Asensio & Pol, 1996). Desde hace tiempo venimos utilizando la técnica de '*clikers*', que permite trabajar estos temas con grupos de personas (familias, escolares, pero también visitantes individuales agrupados) y rastrear una gran cantidad de contenidos con rapidez y economía de medios (Asensio, Asenjo, Castro & Pol, 2014).

Las tareas cuasi-experimentales servirían, en el caso de las aplicaciones tecnológicas, para una cuestión central, que es comprobar la adecuación de los contenidos y por tanto el impacto producido. En este caso estaríamos hablando de una valoración de la didáctica de los contenidos. Realizadas en una fase previa o formativa, estas evaluaciones permitirían reorientar los contenidos en una dirección adecuada para adaptarse más a las capacidades y a las disposiciones de los usuarios.

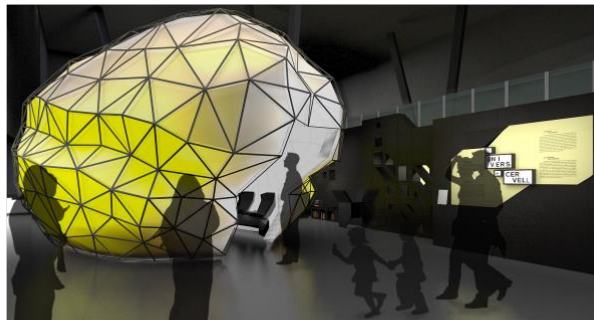
6-. Los *trackings* de los dispositivos como herramienta de evaluación.

Un mundo aún lejano es el que nos permitiría cruzar resultados provenientes de diferentes perspectivas, métodos y herramientas de medida. A todos se nos ha ocurrido que sería interesante contrastar los resultados de *tracking* de las aplicaciones con los resultados de impacto y opinión o de observación. ¿Por qué esto que es tan

evidente no es posible? Porque los trackings están pensados desde una lógica que no es la de la evaluación y raramente son susceptibles de lecturas fiables y unívocas desde otro punto de vista que no sea el del diseño-máquina. Nuestro equipo lo hemos intentado en varias ocasiones y sin embargo solamente una hemos conseguido que finalmente nos den un informe resumido de los trackings, pero era tan general y parcial que apenas daba información.

No parece difícil programar los *trackings* de manera que pudieran servir a intereses de evaluación. Pero para ello, sería necesario un cambio de mentalidad por parte de los diseñadores de las aplicaciones de manera que negociaran los objetivos de los *trackings* con personas de los otros niveles de aplicabilidad del dispositivo.

En la actualidad estamos desarrollando para la Fundación la Caixa una evaluación sobre una exposición temporal que incluye una considerable cantidad de propuestas tecnológicas, desde realidad virtual a interactivos digitales, pasando por un sistema de “singularización de la visita” que consiste en una tarjeta que los visitantes usan para lanzar cada interactivo y que permite a cada ordenador reconocer la tarjeta y guardar personalmente la navegación realizada. La evaluación de la exposición, entre otros aspectos, ha consistido en una tarea de observación de los tiempos y comportamientos desarrollados por 112 visitantes a lo largo de toda la exposición.



Exposición “Talking Brains”, CosmoCaixa, 2017-18.

Este trabajo es la primera vez en la que hemos podido cruzar las diversas medidas de observación con los tracking de las tarjetas. Disponemos también de estudios de ‘aprendibilidad’ de los contenidos (Pol & Asensio, 2017).

En cuanto a su planteamiento general, como veremos a continuación, este estudio sería el trabajo 14 de la investigación, pero lo hemos querido dejar fuera debido a la diferente fuente de financiación.

7-. EDU2014-52675-R: una estrategia de investigación sobre evaluación de APPs.

Nuestra investigación sobre “el análisis evaluativo de las aplicaciones para M-Learning de carácter inclusivo en los espacios patrimoniales” se planteaba rastrear diferentes aproximaciones al estudio de la evaluación de las APPs. Uno de nuestros objetivos era componer un diseño que mezclara diferentes métodos, tipos de estudios y distintas herramientas, agrupándolas en un modelo que, en una primera instancia, coordine conocimientos provenientes de diferentes fuentes y, en una segunda instancia,

podria dar cuenta de la complejidad de la utilidad de las APP en este ámbito de estudio.

Nuestro diseño partía de la consideración, detectada en la revisión, sobre los dos tipos de estudios más habituales en este ámbito: los estudios de indicadores con sujetos expertos y los estudios empíricos con usuarios. En ambos casos se pueden distinguir diferentes niveles de estudios de más genéricos a más particulares: los primeros dedicados a las condiciones generales del uso de la tecnología, de las características generales de las aplicaciones, o de los procesos implicados en su desarrollo y en su uso; y los segundos más centrados en estudios de caso concretos, aterrizados en aplicaciones específicas en lugares determinados.

Nuestra estrategia se basó en varias dimensiones de complementariedad de diferentes tipos de estudio a fin de poder captar una mayor complejidad del fenómeno.

El primer parámetro de complementariedad es el de sumar **diferentes enfoques de análisis**, básicamente dos: los estudios empíricos 'in situ' realizados sobre usuarios reales; y los estudios racionales de carácter ubicuo realizados por sujetos expertos.

El segundo parámetro de complementariedad es el de sumar **diferentes niveles de aproximación**, desde una mirada más genérica a los datos globales de interés en la tecnología en general y en aplicaciones móviles en particular; frente a estudios de caso concretados en aplicaciones en espacios singulares.

El tercer parámetro de complementariedad es el de sumar **diferentes tipos de aplicación** con dos variantes principales, una de *hardware, tablets vs. smart-phones*, y otra de *software*, unas aplicaciones más centradas en el 'modelo de gamificación' y otras más cercanas al 'modelo educativo'.

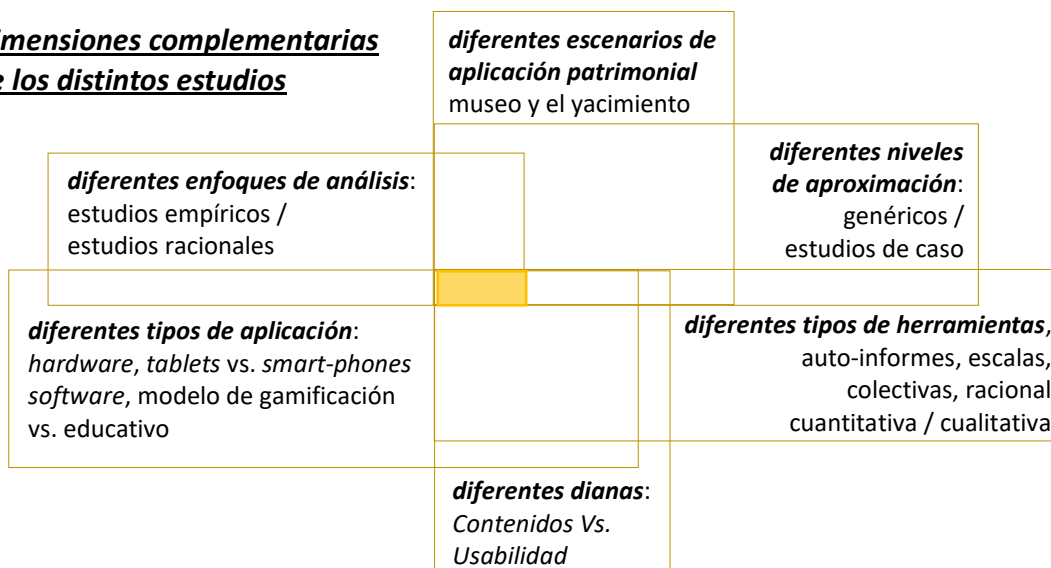
El cuarto parámetro de complementariedad es el de sumar **diferentes tipos de herramientas**, auto-informes (tarea de cuestionarios), escalas de preferencia (tarea de formatos), tareas colectivas negociadas y tarea racional de indicadores (tipo inventario). Todas ellas tienen una parte cuantitativa y otra parte cualitativa.

El quinto parámetro de complementariedad es el de sumar **diferentes escenarios de aplicación patrimonial**, en nuestro caso el museo y el yacimiento, cada uno con sus ventajas y limitaciones.

El sexto parámetro es de complementariedad en el de sumar diferentes tipos de dianas/objetivos de evaluación, con el énfasis en los contenidos o el énfasis en la usabilidad.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

Dimensiones complementarias de los distintos estudios



El cuadro que figura a continuación recoge las tareas realizadas acordes con algunos de los cruces recién comentados. Un conjunto de 13 estudios que empiezan a rastrear la complejidad de estos trabajos. A partir del punto siguiente, describiremos brevemente cada uno de los estudios realizados.

DISEÑO DE LA INVESTIGACIÓN EDU2014-52675-R

	ESTUDIOS EMPÍRICOS CON USUARIOS	ESTUDIOS RACIONALES CON EXPERTOS	ESTUDIOS RACIONALES CON EXPERTOS
ESTUDIOS GENÉRICOS	ESTUDIO 1: Motivaciones y expectativas hacia la Tecnología ESTUDIO 2: Preferencias sobre formatos analógicos y digitales	ESTUDIO 7: Seminario de Test de indicadores de USABILIDAD Revisión y discusión de propuesta de modelo	ESTUDIO 10: Seminario de Test de indicadores de CONTENIDOS Revisión y discusión de propuesta de modelo
ESTUDIOS DE CASO	ESTUDIO 3: Uso de Tablets en el museo: MAN ESTUDIO 4: Uso de Tablets en el yacimiento: Gáldar ESTUDIO 5: Uso de APPs en el museo: Lázaró ESTUDIO 6: Uso de APPs en el yacimiento: Calafell	ESTUDIO 8: Aplicación del modelo experto de análisis de usabilidad: Els Vilars. ESTUDIO 9: Aplicación del modelo experto de análisis de usabilidad: Calafell.	ESTUDIO 11: Aplicación del modelo de análisis de contenidos: Els Vilars. ESTUDIO 12: Aplicación del modelo de análisis de contenidos: Calafell. ESTUDIO 13: Aplicación del modelo de análisis de contenidos: Tarragona.
	DIANA en Usabilidad		DIANA en Contenidos

7.1-. ESTUDIO 1: Motivaciones y expectativas hacia la Tecnología.

Reducir la disposició a una atracció positiva o negativa hacia la tecnologia en base a la generació de referencia, la llamada brecha digital, es un reduccionismo que no nos ayuda a entender el problema. Sin duda, existe una tendencia hacia la brecha digital, pero no es la única fuente que explique la predisposición hacia la tecnología. Se habla a menudo sobre la brecha digital y los datos empíricos a los que hacen referencia son en la inmensa mayoría de casos los datos provenientes de encuestas sobre porcentajes de usuarios por edades. El problema es cómo interpretamos el dato empírico, que en sí mismo es indudable. El corte generacional no es explicativo en sí mismo ya que existen numerosas excepciones, existe una progresiva incorporación de ciertas capas de generaciones de 'boomers' a ciertas tecnologías, o existen sustanciales diferencias entre géneros (por ejemplo, en los datos de gamificación). Indudablemente entre los 'milenials' y las generaciones previas existe una exposición natural a la tecnología que generaciones anteriores no han vivido y es evidente que esto influirá en su disposición hacia la tecnología. Pero, no es menos cierto, que lo que nos estamos encontrando son opiniones y disposiciones muy diversas hacia las tecnologías incluso dentro del mismo corte generacional. Lo mismo que en nuestra generación de 'boomers' encontramos enormes diferencias en el uso de otros dispositivos tecnológicos de la vida cotidiana como la televisión, la lavadora o la *termomix*.

Si es evidente que la brecha generacional no es explicativa, parece que cada vez hay más acuerdo en que tampoco es predictiva ni normativa. Lo que se precisa, en suma, es una investigación más cualitativa y profunda, con herramientas directas de detección de datos empíricos, que profundice en las cuestiones de predisposición hacia la tecnología, y que, de paso, lo haga teniendo en cuenta los modelos de análisis motivacional que desde hace años se vienen utilizando en psicología para otros ámbitos de análisis de las preferencias humanas, como recoge por ejemplo el libro de Richard M. Ryan, *The Oxford Handbook of Human Motivation* (2012).

Una idea central, que profundizaremos con más detalle en las publicaciones específicas de este estudio, consistirá en avanzar en el análisis de lo que podíamos llamar la "experiencia tecnológica patrimonial" (a partir de ahora "*THE*", en sus siglas en inglés *Technological Heritage Experience*). Los modelos de experiencia son deudores de los estudios psicológicos sobre la experiencia autotélica de Csikszentmihalyi (1990, 1998), que en el ámbito de la ergonomía tecnológica dieron lugar a la reflexión y los estudios de Don Norman (1993) y que en nuestro contexto patrimonial ha sido desarrollado como 'experiencia museal', también por varios autores como por ejemplo Falk & Dierking (1992, 2017). Un concepto clave que habrá que seguir desarrollando.

El estudio 1 consiste en el diseño de una tarea original sobre las disposiciones hacia la tecnología, especialmente centrados en el ámbito del patrimonio. Diseñamos una tarea colaborativa y cualitativa, en línea con el tipo de tareas más recientes y modernas en este ámbito de estudio (Kaplan & Patrick, 2016). La base de la tarea explora los dos componentes básicos de los modelos de análisis de la disposición previa: la motivación y la expectativa. Entendiendo la motivación y la expectativa desde una perspectiva general y popular, para que puedan ser perfectamente entendidas por los usuarios: ante la visita a un espacio patrimonial más o menos

concreto, la motivación será sin más una tendencia positiva o negativa hacia el contenido propuesto en torno a ese destino patrimonial; y la expectativa será cualquier previsión de aspectos concretos sobre lo que vamos a encontrar en dicha visita. Un visitante puede encontrarse atraído por exposiciones de geología y puede esperar encontrar en la exposición algunos ejemplares espectaculares en tamaño y peso; respectivamente, estaríamos ante una motivación (M) y una expectativa, ambas positivas (M+ y E+). Pero las motivaciones pueden ser positivas y negativas (por ejemplo, una M – bastante habitual sería “a mí no me gustan las exposiciones que tienen realidad virtual porque no suele aportarme nada”. Y las expectativas ser pueden también positivas y negativas (por ejemplo, una E - bastante habitual es algo así como “si en la exposición hay dispositivos manipulativos digitales, van a ser muy difíciles y no voy a ser capaz de manejarlos”).

La evaluación se realiza en el momento anterior y en el momento posterior a la visita al lugar patrimonial. Es conocido que la evaluación de expectativas y disposiciones previas no puede realizarse con posterioridad a la visita, por tanto, lo que se evalúa en este caso en el post es la idea que qué motivaciones y qué expectativas se han cumplido en la visita (positiva) o no sean cumplido (negativas).

En su sentido más práctico, la tarea consiste en que a un grupo de personas se les proporciona unos *pos-its* de colores diferentes, dos para cada tipo de comentario (M+, M -, E+ y E -). La cantidad de etiquetas es opcional y depende de las restricciones prácticas, es posible usar cantidades mayores en base a la muestra, los tipos de grupo, los espacios, etc. Hemos aplicado esta tarea de manera piloto a un número considerable de grupos hasta disponer de un procedimiento adecuado.



Mural de motivaciones y expectativas antes y después de la visita

La **fase de dimensionamiento** es la primera parte de la tarea, cuando se le pide al sujeto que escriba sus opiniones que tienen que ver con las motivaciones positivas y negativas y las expectativas positivas y negativas. La fase de dimensionamiento

permite sacar a la luz las dimensiones de motivación y expectativas más relevantes. Tras la fase de dimensionamiento, el monitor de la actividad realiza una primera lectura y hace una agrupación de aquellas opiniones más semejantes.

La **fase de escalamiento** es la segunda parte de la tarea. Una vez que las personas han escrito, y pegado en el mural, las dos motivaciones positivas y las dos negativas, y las dos expectativas positivas y las dos negativas, se reparten entre los participantes ocho *gomets* de color verde y otros ocho de color rojo. Se les pide a los participantes que coloquen los verdes sobre aquellas motivaciones y expectativas, tanto positivas como negativas, con las que estén claramente más de acuerdo; y los verdes sobre los que estén más en desacuerdo. La fase de escalamiento permite saber el grado de acuerdo o desacuerdo grupal, es decir, socialmente mediado, y permite sacar un peso para cada una de las dimensiones. El peso puede ser absoluto y relativo. El peso absoluto será el número de gomets recibidos sin distinción del signo positivo y negativo (si una dimensión ha recibido tres gomets verdes / positivos y dos gomets rojos / negativos, tendrá un peso absoluto de cinco). El peso relativo será el resultado de sumar las menciones recibidas y restar las negativas (en el caso anterior el peso relativo sería de $+3-2=+1$). La diferencia entre los pesos absolutos y negativos es notable, ya que el peso absoluto nos está indicando la relevancia de la dimensión con independencia de que existe acuerdo sobre su tendencia. Mientras que el peso relativo sí nos indica esta tendencia. En algunos casos esta distinción se ha mostrado relevante.

La muestra. En el presente estudio la tarea se aplicó a dos grupos de jóvenes, en Madrid y en Barcelona respectivamente. El primer grupo de 18 jóvenes, de edad media de 22 años, todos ellos ya docentes, que además están en formación en un máster de didácticas específicas. El segundo grupo de 5 jóvenes, de edad media de 21 años, todos ellos en formación del grado de turismo de la Universitat de Barcelona.



Aplicación real de la tarea.

Este capítulo pretende dar una visión global del planteamiento general de la investigación y una presentación del diseño de investigación, por tanto, no vamos a presentar resultados, que serán en breve expuestos en publicaciones especializadas. De todos modos, comentaremos dos conclusiones provisionales de esta tarea. Una metodológica, la tarea funciona muy bien, los participantes aportan una gran cantidad

de opiniones, entienden bien la tarea y se sienten satisfechos con su participación; a nivel de diseño, la relación coste / beneficio es más que aceptable; y se cumple el objetivo básico de obtener de los sujetos unas dimensiones de análisis claras y negociadas grupalmente con sus pesos específicos. La segunda, desde el punto de vista del contenido, es que los participantes mostraron una gran diversidad de respuestas, generando una variada amplitud de motivaciones tanto positivas como negativas hacia la tecnología, y de expectativas, también positivas y negativas (lo cual contrasta con otros métodos, como la entrevista y el cuestionario, que suelen ser muy pobres y con contestaciones ambiguas y redundantes). En cuanto al escalamiento, las dimensiones que aparecieron en la tarea no recibieron un gran apoyo, diversificándose mucho los apoyos tanto positivos como negativos, sin que hubiera unas dimensiones que aparecieran claramente como relevantes frente a las demás. Esta escasez de diferencias, podría variar si se aumentara el número de participantes y la variedad de los mismos. Una tarea que nos proponemos completar.

7.2-. ESTUDIO 2: Inclusión de los formatos analógicos y digitales

El estudio 2 consistió en el diseño de una tarea original sobre la aproximación a la preferencia de los formatos tecnológicos. Hablamos de formato para referirnos a un conjunto de restricciones y facilitaciones que presentan los tipos de programas de actuación en el patrimonio. Cada formato manipula de manera diferente al menos cinco parámetros básicos del diseño del programa: el tratamiento de la cultura material; el acceso a materiales complementarios de interpretación; el plan de acción sobre el recorrido de la visita; las interacciones con los visitantes; la opción analógica o digital de interacción, lo cual implica también el tipo de tecnología elegida si la hubiere. Cada formato presenta unas *affordances* determinadas, restricciones y/o facilitaciones que son centrales en el establecimiento de la experiencia de uso. (ver una discusión del concepto de 'formato' en Asensio & Pol, 2008; Asensio y Pol, 2014; mientras que una discusión del concepto de *affordance* puede verse en: Conole & Dyke (2004) *Understanding and using technological affordances*).

Nuestra tarea es una escala de 34 ítems que recogen diferentes formatos de programas en el ámbito del patrimonio, unos tecnológicos y otros analógicos. La tarea del sujeto consiste en expresar el nivel de preferencia ante esos formatos en una escala Likert de seis opciones (1-6), sin opción de término medio. El tipo de formatos son una selección de un fondo de 232 formatos museográficos utilizados por nuestro equipo previamente en anteriores investigaciones (Castro, Pérez & Asensio, 2017), desarrollado para estudiar el grado de inclusión de jóvenes y adolescentes sobre los diferentes formatos museográficos.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.



Ejemplos de los ítems 2 y 23, de los 34 ítems de la tarea de preferencia de formatos

En esta tarea hemos utilizado una muestra de 146 participantes distribuidos en cinco grupos: adolescentes, jóvenes, docentes, profesionales de museos y voluntarios de museos. Estos últimos son los encargados de las visitas educativas en los museos nacionales (los dependientes de la subdirección general de museos estatales del ministerio de cultura).

Tal como dijimos en los puntos 2, además del estudio empírico del usuario, una de las fuentes de progreso en el análisis del uso de las tecnologías es estudiar bien el tipo de *imput*. Normalmente, el *imput* tecnológico de este tipo de formatos no se estudia y no disponemos de tipologías que nos permitan cruzar los formatos y sus variantes con los resultados de comportamiento. Las investigaciones reducen la riqueza del formato al caso individual del APP o Tablet estudiada, perdiendo así una gran cantidad de información. Nuestra tarea permite empezar a realizar un análisis de parámetros de los formatos que ayuden a determinar sus potencialidades en relación a los datos empíricos obtenidos. En este caso, hemos hecho una primera agrupación de los ítems en torno a tres parámetros, a saber: formatos pasivos vs. activos (los formatos pasivos suelen coincidir con los más tradicionales); formatos con objetos vs. sin objetos; y formatos analógicos vs. tecnológicos.

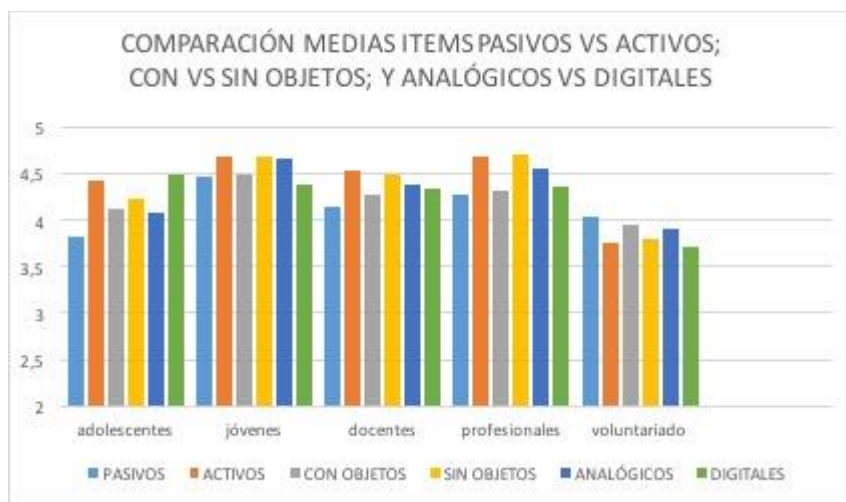
ítems					
+ pasivos	+ activos	Sin objetos	Con objetos	Analógicos	Digital
1, 2, 3, 4, 5, 6, 15, 16, 17, 19, 20, 21, 22, 32	7, 8, 9, 10, 11, 12, 13, 14, 18, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34,	1, 2, 4, 5, 6, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 27, 28,	3, 7, 8, 9, 10, 11, 12, 13, 14, 20, 26, 29, 30, 31, 32, 33, 34,	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 26, 29, 30, 31, 32, 33, 34,	21, 22, 23, 24, 25, 27, 28,

Agrupación de los ítems en torno a parámetros bipolares de caracterización de los formatos.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). *Avaluació d'APP per a un patrimoni inclusiu*. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

Cada ítem satura por tanto en cada parámetro de manera que vendrá finalmente definido por los mismos. La tradicional 'visita guiada' será un formato pasivo, sin objetos y analógico.

Esta tarea da una cantidad muy considerable de resultados. Además, los cruces de variables (por tipo de grupo y por tipo de parámetro, aportan gran riqueza a las discusiones posibles, lo cual implica una interpretación compleja. A continuación, aparece una tabla con los resultados de todos los grupos para todos los tres parámetros bipolares, lo que hace seis dígitos para cada grupo.



Podemos ver que los resultados apuntan diferencias importantes entre unos y otros grupos y parámetros.

El primer dato a destacar es la escasa puntuación media que dan los participantes del grupo de voluntariado a todo tipo de formatos, valorando más los más pasivos. Este es un dato preocupante ya que son los que se encargan de realizar una gran cantidad de las visitas escolares, que es el grupo, los adolescentes, que reclaman precisamente lo contrario, unos programas más activos y participativos.

El segundo dato a destacar, es que los formatos tecnológicos resultan muy bien valorados en general, especialmente por los jóvenes, mientras que los que peor lo valoran son los del voluntariado. Una diferencia notable más entre ambos grupos. Estos resultados serán descritos con detalle en breve en la publicación correspondiente.

7.3-. ESTUDIO 3: Uso de Tablets en el museo: MAN

Ya dijimos que, entre las escasas evaluaciones realizadas, las más frecuentes son las que se dedican a valorar un estudio de caso concreto, bien sea una experiencia escolar sobre un yacimiento, una Tablet a disposición de los visitantes de un museo o un APP que se puedan bajar turistas o escolares en una ciudad histórica.

Nuestro proyecto debía igualmente incluir estudios de caso y evaluaciones de tablets y APPs concretas. Como ya comentamos anteriormente, de la considerable cantidad de aplicaciones que ofrece el mercado decidimos diferenciar entre aplicaciones en el

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

museo o espacio cerrado, frente a aplicaciones para utilizar en un yacimiento o espacio abierto. Además, distinguimos entre aplicaciones para tablets y APPs para teléfonos inteligentes. Finalmente, nuestro criterio era firme en cuanto a seleccionar aplicaciones en uso comercial, no prototipos, sino proyectos ya aplicados, en marcha y que deben cumplir con unos mínimos de mantenimiento y sostenibilidad, ninguna de las aplicaciones evaluadas era una versión experimental, ni siquiera beta.

Nuestros estudios de caso se realizaron también con una herramienta clásica, el cuestionario en su versión más típica de informe detallado auto-informado.

El estudio 3, por tanto, consistió en el diseño de una tarea original, pero basada en los guiones de cuestionario habituales en este tipo de estudios (ver por ejemplo una interesante reflexión en Economou & Meintani, 2011).

La tarea incluye 19 preguntas cualitativas, de las cuales 8 incluyen además una estimación cuantitativa directa escalar (tipo Likert) y otras 4 tienen una estimación cuantitativa ordinal, el resto permiten análisis cuantitativos categoriales.

La tablet seleccionada se correspondió con la utilizada en el Museo Arqueológico Nacional, realizada y mantenida por una empresa externa (GVAM), que se adquiere en el hall de acceso sobre un suplemento de precio en la entrada al museo. El primer dato a analizar sería cómo se anima su uso y cuál es el porcentaje de uso y los perfiles de sus usuarios. Informalmente diremos que su publicitación es prácticamente nula y no se recoge ninguna información sobre los

Este estudio presenta una muestra de 22 docentes jóvenes y sus resultados iniciales no son muy favorables a la aplicación. La opinión generalizada es de que se trata de una herramienta totalmente centrada en los contenidos, pesada y lenta, sin ningún atractivo para los escolares, adolescentes o no iniciados, con contenidos muy técnicos y monótonamente presentados. Las notas cuantitativas son las más bajas que hemos encontrado para este tipo de aplicaciones entre los estudios realizados en esta investigación. En cuanto a las características del propio formato también se describen muchas limitaciones de accesibilidad y usabilidad, aunque la propia simpleza de la aplicación facilita en cierto modo que disminuyan los problemas e este sentido.



Montaje del MAN puesto de ejemplo por los participantes como uno de los manipulativos (en este caso tecnológico) potencialmente atractivo pero claramente desaprovechado.

7.4-. ESTUDIO 4: Uso de Tablets en el yacimiento: Gáldar

El estudio de las tablets del Museo y Parque Arqueológico Cueva Pintada son otro modelo radicalmente diferente, primero porque están pensadas para acompañar la visita por el yacimiento, pero sobre todo porque incluyen en el propio programa un sistema de evaluación tanto de la experiencia realizada como del propio programa. Si bien la “accesibilidad física” al patrimonio es fundamental, los espacios de presentación del patrimonio deben ser inclusivos con absolutamente todos los visitantes reales y potenciales y para ello debe empezar a reflexionar sobre por qué determinados colectivos no acuden a nuestras instituciones y reflexionar sobre si realmente les estamos ofreciendo algo con lo que se sientan identificados, que despierte su interés y de lo que deseen formar parte. Uno de los colectivos del cual somos más dolorosamente conscientes de que no conseguimos que se sientan incluidos en nuestras propuestas expositivas y educativas (sin la “coacción” de padres y profesores) es el colectivo de los adolescentes. Lo que algunos autores llaman ya descarnadamente el museo ‘intimidatorio’ (van Boxtel et al., 2016); o lo que Nancy Proctor (2011) proponía sobre la tecnología como facilitador de la experiencia en el museo.

Nuestro equipo de investigación lleva colaborando con el Museo y Parque Arqueológico Cueva Pintada (Gáldar, Gran Canaria) desde su inauguración en 2006. El motivo inicial fueron los estudios de público (Asensio, 2017), de los que llevamos ya once años acumulando los perfiles los impactos y la opinión. Pero, además, se ha venido haciendo una labor de evaluación sistemática de exposiciones permanente y temporales, y de aspectos de uso y comprensión y aprendizaje en los programas públicos y educativos. Como consecuencia de estas acciones de evaluación se ha realizado una asesoría permanente sobre temas de formación y de propuestas de nuevos espacios expositivos y de programas. Todo ello desde una estrecha labor de colaboración con los técnicos del museo y del cabildo, donde recientemente se han ampliado ciertos estudios a otros centros de la red de museos.

Entre las propuestas realizadas, se diseñó un espacio nuevo, el i-Lab (Asenjo et al., 2017), como un entorno de aprendizaje profundo mediado por las Tecnologías de la Información la Comunicación y la Participación (TICP). Un espacio de trabajo diseñado especialmente para el público de secundaria: el Laboratorio Interactivo de Aprendizaje de la Historia. Por otro lado, el i-Lab supone un programa educativo complejo, dirigido a la elaboración y reelaboración profunda de conocimientos vinculados a Cueva Pintada, así como al desarrollo de habilidades transversales, como el manejo de mapas, el trabajo en equipo y el uso de tecnologías para el aprendizaje. Finalmente, el programa también incluye el desarrollo de aspectos emocionales y motivacionales vinculados tanto hacia los contenidos patrimoniales como hacia el propio proceso de aprendizaje. De este modo, el objetivo final es crear un espacio especialmente diseñado para que los adolescentes hagan suyo tanto el patrimonio como la institución y puedan trabajar los contenidos de forma autónoma, atractiva, amable, cercana y profunda.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.



Alumnado de secundaria trabajando en el iLab del Museo y Parque Arqueológico Cueva Pintada.

Durante el diseño de lo que sería el “espacio tecnológico” del i-Lab teníamos claro que la herramienta tecnológica debía acompañar a los participantes durante el recorrido por los espacios expositivos, para que sirviese de apoyo, andamiaje y dinamizador del contacto con la cultura material y con los contenidos del mensaje expositivo, así como durante su estancia en el espacio físico del i-Lab, para que sirviese de igual modo al marcaje y a la reelaboración de los contenidos vistos durante la visita. Por tanto, la herramienta utilizada debía ser portátil y favorecer el *mobile learning* (Handal, 2016) o *ubiquitous learning* (Bruce, 2009), es decir, que sean fácil mente transportable, personalizables, usables, etc. (Sharples, Taylor y Vavoula, 2007). Sobre esta premisa nos decantamos por el formato del iPad mini, por su amplia pantalla que permite trabajar cómodamente los contenidos y su ligereza que lo hace adecuado para transportarlo durante un periodo de tiempo prolongado. A petición del profesorado que participó en contactos previos, el *software* utilizado, dado que el objetivo es que aprendieran a usar los recursos tecnológicos a su alcance con fines de aprendizaje / alfabetización tecnológica, decidimos emplear aplicaciones gratuitas de libre acceso en Internet, así como aplicaciones básicas del propio iPad. El programa educativo del i-Lab es un programa de *Deep Learning* o aprendizaje profundo (Herman & Linn, 2013; Hermida, 2015). Es decir, un programa que cree las condiciones necesarias para que se produzcan aprendizajes que permanezcan durante un periodo de tiempo prolongado y que pueda ser aplicado en nuevas situaciones y contextos.

A lo largo de esas cuatro horas los estudiantes tendrán que participar en un total de tres fases que componen el programa: Fase 1: Acogida e introducción: presentación de la actividad y de los contenidos y evaluación previa y plan de trabajo; Fase 2: Visita al Museo y Parque Arqueológico: realizando las actividades propuestas; y Fase 3 Reflexión y cierre: revisión del trabajo realizado, envío de los materiales trabajados a sus centros, evaluación sumativa, autoevaluación y despedida. La evaluación se lleva a cabo con *clickers*, mandos de votación remota (Asensio, Rodríguez y Sáenz, 2012). A continuación, los grupos, equipados con su iPad, realizarán el recorrido de forma autónoma, pero pautada por unas líneas de actuación alojadas en el iPad.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.



Alumnado trabajando con las Tablets en el Museo y Parque arqueológico Cueva Pintada.

Los primeros resultados se basan en dos estudios, con estudiantes de 3º y 4º de la ESO del IES Alonso Quesada. En general, el programa contó con una gran aceptación por parte de los estudiantes y de los profesores (el uso de las iPads obtuvo una puntuación de 9,4 sobre 10). Además, también se registró una alta satisfacción respecto al patrimonio material (colección del museo y yacimiento) y respecto a los recursos museográficos (museo y audiovisuales), que obtuvieron puntuaciones entre 8,75 y 9,2 sobre 10. Si comparamos grupos que realizan la visita guiada frente a los grupos que realizan la visita con la Tablet, todos los grupos presentan una satisfacción bastante alta en la valoración general de la visita, pero el grupo que participó en el i-Lab presenta una satisfacción significativamente mayor respecto a la colección, el museo y los audiovisuales. Esto nos indica que el fuerte efecto atractor que genera el programa del i-Lab, puede estar transfiriéndose a los elementos museográficos concretos. Finalmente, los comentarios expresados por los participantes fueron, en general, muy positivos (Ej. “Más entretenido que la visita guiada”; “se nos pasó el tiempo volando”; etc.).

7.5-. ESTUDIO 5: Uso de APPs en el museo: Lázaro

El Museo Lázaro Galdiano de Madrid, dispone de una APP muy interesante, desarrollada por el Departamento de Ingeniería del Software e Inteligencia Artificial Universidad Complutense de Madrid. La aplicación, denominada ‘Enigma Galdiano’, está diseñada tomando como referente los ‘serious games’ y consiste en un recorrido por las salas del museo resolviendo una serie de preguntas sobre las obras a partir de un sistema de realidad aumentada. Una descripción más técnica y unos primeros datos de evaluación puede verse en Camps et al., 2017).

El estudio 5 consiste en el diseño de una tarea original para evaluar impacto y opinión sobre la aplicación ‘Enigma Galdiano’. La tarea consiste en un cuestionario de autoinforme de 43 preguntas, 20 abiertas y 23 con forma de escala Lickert.

Una versión de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). *Avaluació d'APP per a un patrimoni inclusiu*. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueología & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.



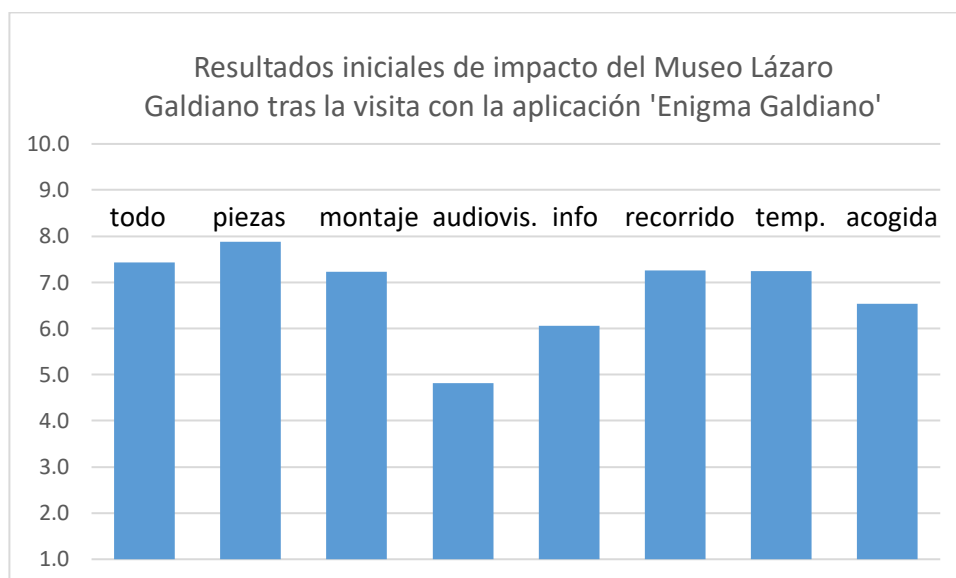
Piense en la visita que acaba de realizar

Marque en los círculos con una (x)

¿Cuánto le ha gustado...?	    									
	Nada	Muy Poco	Poco	Bastante	Mucho	Muchísimo				
	1	2	3	4	5	6	7	8	9	10
En general, todo el Museo	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
En general, las piezas expuestas	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
La disposición y presentación, el montaje	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Detalle del protocolo de la aplicación del Museo Lázaro Galdiano.

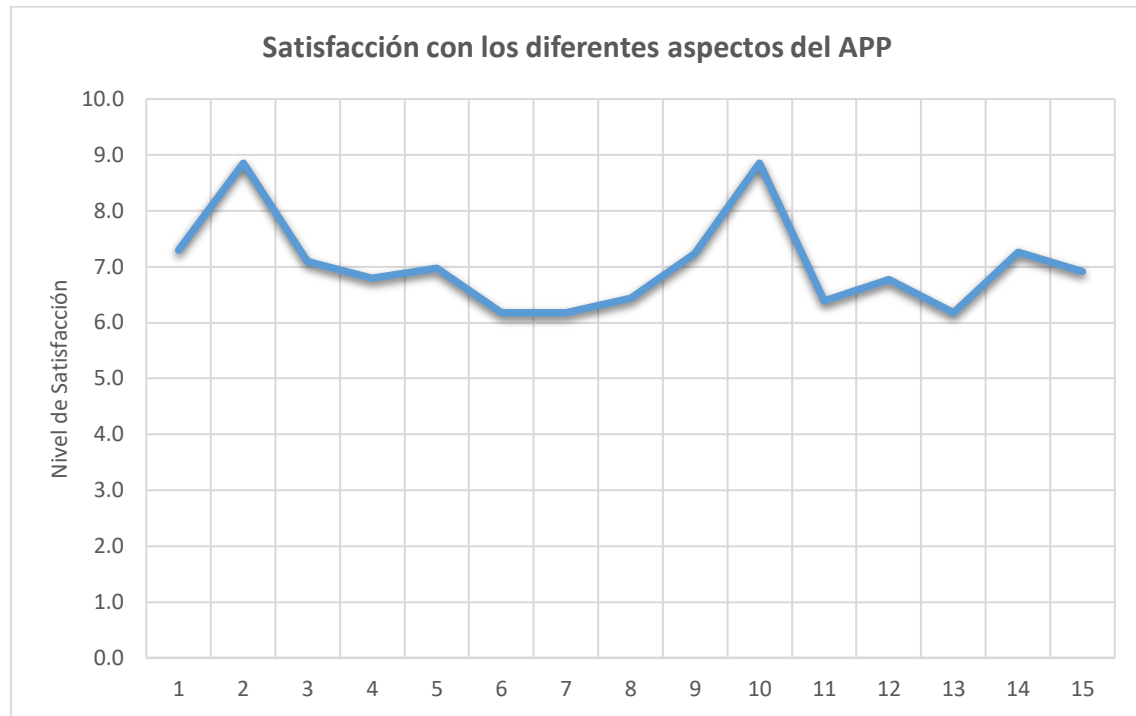
La muestra de un total de 35 personas jóvenes iniciadas. Todas ellas eran jóvenes entre 20 y 32, distribuidos en dos grupos: 15 pertenecen a un grupo de maestros y licenciados dedicados a la docencia; y 20 son jóvenes, estudiantes o recién licenciados y graduados en psicología. Todos ellos eran iniciados e interesados específicamente en temas de patrimonio y nuevas tecnologías; fueron invitados explícitamente a participar en esta experiencia, y todos ellos tenían experiencia previa en haber realizado alguna visita al patrimonio con algún tipo de dispositivo móvil y todos ellos habían manejado varias aplicaciones móviles de este tipo en algún otro contexto, educativo y/o turístico. Los resultados iniciales indican, no solamente una gran opinión del uso de la aplicación, sino una buena opinión de la mayor parte de las dimensiones de opinión generales del museo.



Dos tercios de la muestra reconocen explícitamente que el museo es muy interesante mientras que un tercio lo consideran clásico en cuanto a su planteamiento. La mayoría destacan el atractivo de sus colecciones y del montaje.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

Pero los resultados más interesantes son los que se refieren a la satisfacción del uso de la aplicación y de sus diferentes partes. La gráfica muestra los resultados de los aspectos a evaluar que se especifican en la tabla de abajo.



	Aspectos a evaluar	Medias de Satisfacción		
1	El planteamiento general de la aplicación	7,3	7,5	7,1
2	La idea de usar un juego para la visita	8,9	8,5	9,1
3	La historia / argumento del juego	7,1	7,2	7,0
4	Los personajes y su trama	6,8	7,1	6,6
5	La visita con el juego te ha parecido divertida	7,0	6,7	7,2
6	La duración del juego	6,2	4,9	7,2
7	La visita con el juego es fácil de seguir	6,2	5,0	7,0
8	Las pistas han sido fáciles y eficaces	6,4	5,3	7,3
9	El juego	7,2	8,0	7,4
10	La idea de usar un juego para la visita	8,9	8,3	9,1
11	El mapa	6,4	5,6	6,9
12	La estructura de las islas	6,8	6,7	6,8
13	La relación del mapa con las obras	6,2	5,5	6,7
14	El loro	7,3	6,9	7,6
15	Los mini-juegos	6,9	6,7	7,1
		Media total	Media docent	Media iniciad.

Dos aspectos complementarios muy interesantes resultaron los referidos a la estimación de para qué tipos e públicos es más adecuada la aplicación y la opinión sobre el carácter lúdico/educativo de la aplicación.

Hay un acuerdo bastante generalizado de que el APP es apropiado para familias con niños pequeños o para grupos de niños pequeños, en general, niños de primaria, entre los 8 y los 12 años. Hay menos acuerdo en cuanto a su uso en secundaria y empieza a ser mayoritaria la opinión de que es menos adecuada para jóvenes o adultos, especialmente interesados en los contenidos temáticos.

Por su parte el carácter lúdico vs. educativo de la aplicación viene dado por el convencimiento en ambos grupos por igual de que la aplicación es muy lúdica (media de 4,4 sobre 5, con una pequeña diferencia 4,7 en los docentes frente a 4,2 en los jóvenes iniciados), mientras que la aplicación resulta muy poco educativa (la media es igual de baja, 2,8 sobre 5 en los dos grupos).

7.6-. ESTUDIO 6: Uso de APPs en el yacimiento: Calafell

La ciudadela ibérica de Calafell (Tarragona), de 2.500 años de antigüedad, es un yacimiento arqueológico que tiene la peculiaridad de estar reconstruido en parte, y que cuenta con una contextualización de piezas y bienes muebles que permiten interpretar adecuadamente muchos de los espacios. El reto de la reconstrucción, y por tanto, la elección de determinados momentos y formas, muchas de ellas basadas en la arqueología experimental, aporta a las APPs una excelente oportunidad para realizar una aporte significativo sobre las diferentes variantes de interpretación y sobre los criterios utilizados. La *app*, denominada Calafell Open Air Museum, utiliza la tecnología de la realidad aumentada y las posibilidades de la realidad virtual para mostrar cómo era esta población en diferentes épocas de su historia (siglos VI-III aC). La herramienta es gratuita –se puede descargar en iTunes y Google Play– y ha sido desarrollada por la empresa DIGIT con la colaboración del Grupo de Didáctica de la Historia, la Geografía y otras Ciencias Sociales(DHiGeCS).

Nuestro proyecto de investigación siempre tuvo como objetivo la evaluación de esta aplicación. El estudio 6 consiste en la adaptación de la tarea del estudio 5, a la aplicación de la ciudadela ibérica de Calafell. Usar la misma tarea tiene el sentido de poder facilitar las comparaciones con tareas equiparables.

La tarea consiste en un cuestionario de 41 preguntas: 3 iniciales antes de la visita, 11 preguntas abiertas y 27 preguntas cerradas escaladas tipo Lickert, de ellas las 10 primeras recogen aspectos generales de la visita, mientras que las 17 restantes recogen los aspectos concretos referidos a la aplicación. El procedimiento de aplicación es siempre supervisado, con una introducción inicial en la que se explica la tarea y os objetos, se presenta el procedimiento, se contestan todas las dudas, se responsabiliza sobre la importancia de la actividad y se oferta un obsequio al final de la visita. También se ofrece la posibilidad de dar feed-back sobre los resultados, tanto a la institución como a los participantes interesados. Todos los cuestionarios son anónimos.

Detalle del protocolo de evaluación de la aplicación de la visita a la ciudadela Ibérica de Calafell

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.



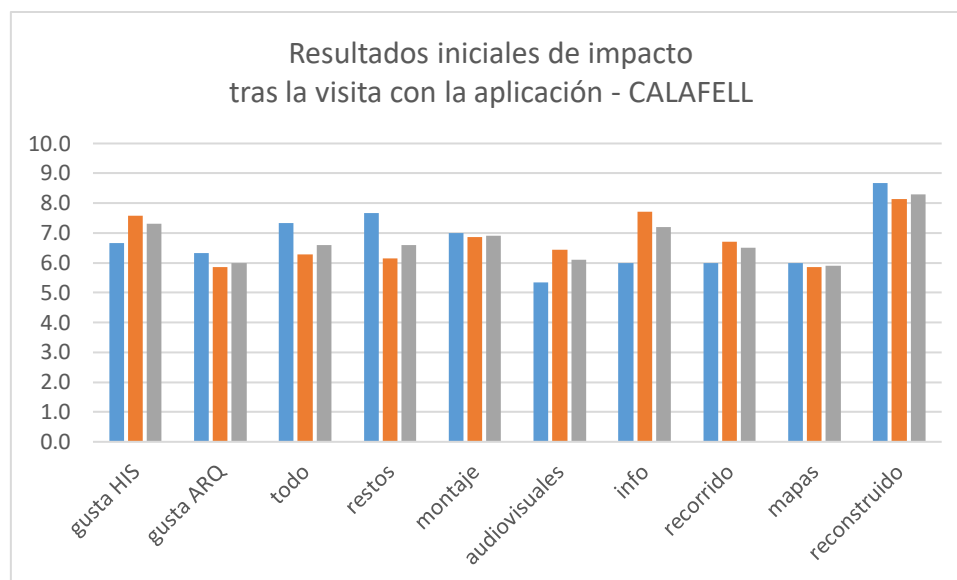

Piensa en la aplicación que ha utilizado durante la visita que acabas de realizar

Marque en los círculos con una (x)

¿Cuánto te ha gustado...?	Nada 1	Muy Poco 2	Poco 3	Poco 4	Bastante 5	Bastante 6	Mucho 7	Mucho 8	Muchísimo 9	Muchísimo 10
El planteamiento general de la aplicación	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cómo se guía y desarrolla la visita	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
La historia / argumento que va contando	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
El tratamiento de las personas	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
La visita te ha parecido divertida	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

La muestra del estudio de Calafell es por ahora exigua, disponemos de resultados de 3 profesores y de siete alumnos de secundaria. Todos ellos hicieron la visita con la aplicación, y rellenaron el cuestionario después de la visita.

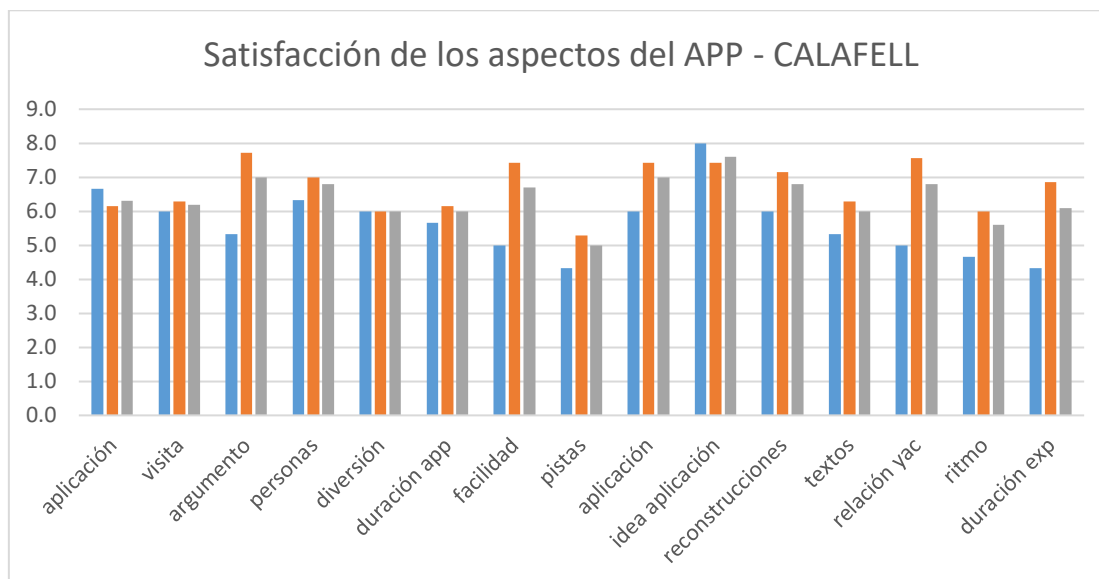
Los resultados iniciales muestran una buena opinión del yacimiento y sus dimensiones tras la visita con el APP.



Los resultados en azul son los globales mientras que los naranjas son los de los profesores y los grises los de los adolescentes. Pueden observarse diferencias notables entre los diferentes aspectos, uno de los más destacables es la buena opinión de los tres grupos sobre las reconstrucciones realizadas.

Pero, tal como ocurría en el estudio 5, los resultados más interesantes de este estudio 6 se deben a los referidos a los aspectos evaluados de la propia APP. Los grupos de colores son los mismos que en la gráfica anterior.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.



Los resultados muestran que mientras la opinión general sobre la aplicación y sus aspectos más generales, como el argumento y las reconstrucciones, es más positiva, algunas de las cuestiones puntuales como las pistas que ofrece, su duración o el ritmo son menos valorados.

7.7.- ESTUDIOS 7, 8 y 9: Seminario de Test de indicadores de USABILIDAD: revisión y discusión de propuesta de modelo.

El estudio 7 inicia otra línea de investigación, la de un análisis racional, mediante la técnica de indicadores, valorados en principio por personas expertas o al menos iniciadas.

El objetivo de la tarea consiste en conformar un modelo de evaluación de APPs a partir de la selección de un conjunto de indicadores que den cuenta de la complejidad de los aspectos a evaluar. Aquellos que se consideran significativos desde un determinado modelo teórico de referencia, para lo cual es necesario desarrollar en primer lugar un estudio de revisión en el que se perfilan para su comparación a poder ser la mayoría de los modelos significativos propuestos por los especialistas.

El estudio se planteó en dos fases, sobre la base de dos metodologías de trabajo cualitativo colaborativo: el 'Taller de Futuro' y las 'Matrices de Reflexión' (Francés, 2016). **La primera fase es el 'Taller de Futuro'**, que permite la revisión en profundidad de un tema, con una división de tareas (en este caso la revisión bibliográfica de los modelos de evaluación de APPs), el desarrollo progresivo de una puesta en común, y la negociación colaborativa de una propuesta de nuevo modelo de indicadores. **La segunda fase es las 'Matrices de Reflexión'**, que ordena la discusión sobre la aplicación exploratoria de un modelo, la aplicación del modelo a uno o más estudios de caso, el ajuste del modelo en un análisis confirmatorio, y la propuesta final del modelo. En nuestro caso, se aplicaría a dos estudios de caso, que constituirían los estudios 8 y 9.

En la medida que ya hemos dicho que este estudio de las APPs en patrimonio es muy interdisciplinar, hay que revisar bibliografías provenientes de muy diferentes campos de estudio y con todos ellos tratar de fundir las variables en un modelo común que profile los criterios necesarios. Una vez tenidos los criterios se aterrizan éstos en indicadores que deben cumplir unas condiciones de aplicabilidad, claridad, etc. Para poder ser aplicados de manera unívoca. Por ejemplo, Boyle et al. (2016) distinguen siete criterios sobre los que construir los indicadores: adquisición de conocimiento, adquisición de habilidades, perceptuales y cognitivos, fisiológicos, afectivos, cambio de comportamiento, y habilidades sociales. Mientras que por su parte Lee & Cherner (2015) proponen 24 criterios asociados en tres bloques: instrucción, diseño y involucramiento. Economou & Meintani (2011) se basan en cinco bloques: tipo de aplicación, usos en relación al museo, interacción usuario-contenido, integración perspectivas múltiples, e interacción social (del conjunto de artículos que proponen diversos modelos podríamos destacar por diversos motivos los siguientes: Buckler, 2012; de Freitas, 2010; Walker, 2010; O'Brien & Toms, 2008; Reeves, & Hedberg, 2003).

La estrategia sería la de ser capaces de construir un modelo de indicadores que posteriormente sea estilizado mediante procedimientos estadísticos (dejando aquellos indicadores que cumplen con características de validez, fiabilidad, discriminación, factorialización, etc.).

El 'taller de futuro' se ha desarrollado durante más de dos meses sobre la revisión y discusión de los modelos y la extracción de criterios y la producción de indicadores; se han discutido estos indicadores hasta elegir un conjunto con los que construir la tarea. Este estudio está aún en marcha.

El ESTUDIO 8, la aplicación del modelo experto de análisis de usabilidad al estudio de caso del yacimiento de la ciudadela Ibérica de 'Els Vilars', ha sido iniciado con un modelo tentativo, que ha servido para ir chequeando la viabilidad de los criterios y los indicadores. Se realizó una visita al yacimiento con cuatro expertos que permite dos tipos de análisis: uno acumulativo que afecta a la evaluación de los diferentes aspectos de usabilidad de la aplicación; y uno interjueces, que permite analizar la coherencia estadística del modelo a través de la ejecución en los cuatro jueces. La visita fue desarrollada con el APP "Vilars d'Arbeca", de la 'Fortalesa dels Vilars' para teléfonos móviles de sistemas iOS y Android, desarrollado por la empresa *Develoop Software*, con la colaboración del *Grup d'Investigació Prehistòrica de la UdL*, y la conceptualización y el diseño del grupo *DHiGeCS, Didàctica de la Història, la Geografia i altres Ciències Socials*, de la UB. El APP permite recorrer la fortificación e ir accediendo a información e imágenes con restituciones interpretativas y videos de arqueólogos (Natàlia Alonso, Emili Junyent, Joan López, ...). El APP incorpora además unas espectaculares imágenes virtuales de Josep R. Casals.

El APP de Els Vilars aporta una información fundamental para comprender el yacimiento, a nivel de contenidos el APP ha sido analizada en el estudio 11 (descrito anteriormente). A nivel de usabilidad, aunque aún no estén terminados los análisis, podemos avanzar que el APP muestra un conjunto de problemas similares a los detectados en otras APP equivalentes, de una parte, un conjunto de limitaciones técnicas (la geolocalización, de descarga, etc.); de otra, un conjunto claramente

acusado de ubicación de los recorridos y de localización de determinadas estructuras, que confunden considerablemente la movilidad por el yacimiento; a nivel de comprensión, el APP aporta un discurso con un considerable nivel de encriptación, conceptualmente abstracto, que no parece accesible cognitivamente a cualquier nivel de instrucción.

El ESTUDIO 9, la aplicación del modelo experto de análisis de usabilidad al estudio de caso del yacimiento de la ciudadela Ibérica de 'Calafell', se realizará en el momento en que cerremos el modelo de indicadores. Por el momento, disponemos de la aplicación (que ha sido utilizado para el estudio 6 con visitantes reales en el yacimiento). Igualmente el APP ha sido analizada en cuanto a la evaluación de contenidos en el estudio 12 (anteriormente descrito).

El APP de la ciudadela ibérica de Calafell se denomina "CALAFELL OPEN AIR MUSEU". La aplicación utiliza el software de *Metaio*, una empresa de Realidad Aumentada reconocida internacionalmente, con sede central en Múnich, que dispone de un navegador para Android y para iOS. La aplicación ha sido desarrollada por el equipo del grupo DHIGECs de la Universidad de Barcelona, conjuntamente con la empresa DIGIT, que ha realizado la producción. El desarrollo iconográfico fue un trabajo conjunto del equipo de DHIGECs con Guillem H. Pongilupi y Josep Ramón Casals. Los resultados de la evaluación desde el punto de vista de la usabilidad, igual que el estudio anterior, permitirá los mismos dos niveles de análisis, acumulativo (que aportará resultados sobre el análisis concreto de los niveles de uso de la aplicación) e interjueces. En el caso de estos últimos, serán acumulables a los anteriores de modo que sirvan para analizar las propiedades psicométricas del modelo de análisis y criterios y el test de indicadores.

Una vez desarrollados los estudios 8 y 9, se desarrollaría una sesión denominada 'matrices de reflexión', que permite discutir los resultados obtenidos en la aplicación del modelo a los dos estudios de caso (Els Vilars y Calafell), analizar sus propiedades psicométricas y discutir el ajuste final del modelo a nivel teórico. Una vez realizado este paso, el modelo quedaría dispuesto para su aplicación a nuevos casos.

Finalmente, recordar, tal como venimos comentando, que los estudios dedicados a los análisis de contenidos, estudios 10, 11, 12 y 13, ya han sido descritos anteriormente.

8-. Algunas conclusiones provisionales del proyecto EDU2014-52675-R.

Los enfoques o mentalidades implicados en la demanda, construcción y gestión de APPs, que vienen de ámbitos de conocimiento muy diferentes, están cada vez más sintiendo la necesidad de colaboración interdisciplinar, pero aún se está muy lejos de incluirla en los procesos habituales de trabajo. La revisión muestra, en primer lugar, una enorme atomización de enfoques, métodos y herramientas, con escasa regularidad y acumulación del conocimiento. Hay que trabajar sobre la confluencia de estos enfoques, tratando de acercar y complementar los criterios que provienen de las diferentes disciplinas, de manera que construyamos herramientas más completas y comprensivas.

Desde un punto de vista metodológico, creemos indispensable el cruce de al menos tres métodos complementarios: Los estudios racionales con expertos, los estudios

empíricos con usuarios y los estudios de observación y tracking (también con una base empírica, pero que incorporan los datos de tracking de la tecnología). Se necesitan estudios que complementen datos que provengan de métodos diferentes porque captan realidades complementarias. Sería conveniente disponer de un conjunto de tareas específicas, que tengamos el convencimiento de que cumplen con las con el doble rasero teórico (sobre los constructos y criterios) y empírico (sobre las propiedades estadísticas adecuadas de la medida), suficientemente flexibles para que puedan ser adaptadas a los casos concretos.

En ese sentido, las tareas propuestas en nuestra investigación empiezas a construirse con estos criterios y trabajan en esa dirección. En resumen, la mayoría de las tareas propuestas han dado juego desde el punto de vista de sacar información relevante sobre los problemas teóricos principales, y además se han comportado adecuadamente desde el punto de vista estadístico. Sin embargo, estamos aún lejos de conseguir herramientas suficientemente amplias, significativas y flexibles.

Desde el punto de vista de los resultados de evaluación, las APPs están muy lejos de ser herramientas culturales eficientes. Los datos muestran varias evidencias bastante repetidas en torno a unas tendencias incuestionables.

Una primera es la dificultad tecnológica. Las condiciones técnicas y las prescripciones tecnológicas de las aplicaciones suelen ir al límite de las capacidades de las propias plataformas que las soportan, de las propias aplicaciones y de los contextos de uso, incluidos los propios dispositivos de los usuarios. No es este el lugar para desplegar un repertorio de limitaciones técnicas, pero recordemos que, solamente en las evaluaciones realizadas en esta investigación, han salido una considerable cantidad de ellas. Una parte tienen que ver con limitaciones de las propias plataformas, ya que muchas empresas, no se sabe si presionados por el mercado o presas de su propia calentura, llevan al límite las capacidades de las tecnologías que muchas veces no se adaptan a la realidad que se pretende desarrollar en casos concretos. Otra parte tienen que ver con las limitaciones de las características de uso de las instituciones o de los propios usuarios, que no disponen de suficientes capacidades de cobertura, de descarga, etc., lo cual genera problemas prácticos con independencia de la calidad o de la disponibilidad de la aplicación.

Existe una mentalidad muy extendida entre los gestores (que los productores aprovechan con astucia) sobre la idea de que la tecnología es un objeto de deseo para el usuario. Con independencia de que esta es una idea compleja que precisaría una discusión más amplia, nos gustaría destacar que la expectativa, la motivación y la opinión del usuario hacia la tecnología no siempre son positivas. Por el contrario, en muchas ocasiones los usuarios muestran opiniones muy críticas y son capaces de discriminar entre aspectos muy finos sobre la adecuación de la tecnología al uso propuesto.

Si bien un planteamiento adecuado de los mediadores tecnológicos resulta atractivo para conjuntos importantes de usuarios, todos ellos valoran sobremanera que esté diseñada desde el punto de vista de la facilidad de uso. Numerosos autores han denunciado la falta de sensibilidad hacia las demandas del diseño inteligente, desde el diseño del usuario, también llamada UX, experiencia desde el usuario. Las opiniones de los visitantes son igualmente críticas con una gran cantidad de problemas que

presentan las aplicaciones y que tienen que ver con temas de accesibilidad, usabilidad o navegabilidad. Igualmente, son sensibles y aprecian los guiños que influyen las aplicaciones y que sirven para resolver o al menos mejorar este tipo de problemas, como las pistas, las marcas, los bypass, etc. Incluso las propuestas tecnológicas más atractivas pueden presentar inconvenientes, como las gafas de realidad aumentada (Asensio et al., 2013) o los museos virtuales (López-Menchero et al, en prensa). Desde el punto de vista de los contenidos que se implementan en las aplicaciones, la mayoría de las que hemos revisado en la investigación tienen dos problemas. El primero es una falta de rigor disciplinar, histórico o arqueológico. El segundo es la falta de adecuación didáctica. Muchas aplicaciones reproducen información superficial recopilada de fuentes generalistas que reproducen errores históricos o que al menos que no recogen un planteamiento actualizado de los contenidos, sin un proceso actualizado de documentación e interpretación. Un dato importante de cara a la evaluación es que, salvo en raras ocasiones, este primer punto no es transparente a los usuarios, lo cual es entendible. Sin embargo, sí puede observarse de manera sistemática en los informes cualitativos de los expertos y en los test de indicadores. Además, todas las aplicaciones revisadas, incluso las que hemos seleccionado en la investigación para su estudio, que tenían que cumplir con un nivel de rigor disciplinar, comenten errores importantes desde el punto de vista didáctico, es decir, desde el punto de vista de la adecuación de los contenidos a la situación de transmisión del conocimiento. Recuérdese que estamos hablando de resultados de evaluación, no desde nuestra propia opinión (por compartida que sea con la de los usuarios). Son las personas que han utilizado las aplicaciones las que nos dicen que en su opinión hay demasiada información, que la aplicación dura demasiado, con demasiadas paradas, con mucha información gráfica o con un exceso de actividades y propuestas. Finalmente, la mayoría de las aplicaciones ofrecen de manera explícita una experiencia atractiva y divertida, incluso lúdica. Sin embargo, las opiniones de los usuarios son mucho más críticas y mucho más finas y discriminativas. Por ejemplo, pueden valorar positivamente el carácter lúdico, pero criticar al mismo tiempo su escasa potencialidad educativa. O incluso son capaces de valorar de manera medida la cantidad y el tratamiento de la información que debería incluirse en la aplicación si se quiere que tenga un carácter educativo. Un aspecto muy a destacar en el ámbito del patrimonio en el que nos situamos, es la consideración que hacen los visitantes sobre la relación con los objetos y la cultura material, valorando aquellas aplicaciones que consiguen transmitir ideas, representaciones y reflexiones sobre las colecciones, o las obras y artefactos y, a través de ellos, sobre las culturas de referencia. Estas son solo algunas de las tendencias de opinión que estamos encontrando en nuestras evaluaciones y que esta investigación nos ayudará a ordenar y dimensionar. Esperemos.

References.

Asenjo, E., Asensio, M., Castro, Y., Pol, E., Reyes, M.M., Rodríguez, M., Rodríguez, C.G. & Sáenz, J.I. (2017). i-Lab: un espacio de aprendizaje profundo de la arqueología en

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

- el Museo y Parque Arqueológico Cueva Pintada. *Arkeologi Museoa: Los Cuadernos del Arkeologi*, 8, pp. 137-156.
- Asensio, M. (2015). El aprendizaje natural, la mejor vía de acercarse al patrimonio. *REVISTA EDUCATIO SIGLO XXI*, 33, 1, 55-82.
- Asensio, M. (2017). Como el agua que fluye: formar para lo informal. *Información Psicológica*, 113, pp. 70-84
- Asensio, M. (2017). ¿Hacia dónde vamos? Los estudios de público, un desafío para el futuro de los museos. En: Bialogorski, M. & Reca, M.M. (Eds.) *Museos y Visitantes. Ensayo sobre estudios de público en Argentina*. pp. 13-25.
- Asensio, M. y Asenjo, E. (Eds.) (2011): *Lazos de Luz Azul: Museos y Tecnologías 1, 2 y 3.0*. Barcelona: UOC.
- Asensio, M., Asenjo, E., Castro, Y. & Pol, E. (2014). Evaluación implicativa: hacia una visión generativa y participativa en la gestión de audiencias. En: Arrieta, I. (Ed.) *LA SOCIEDAD ANTE LOS MUSEOS. Públicos, usuarios y comunidades locales*. Bilbao: EHU-UPV. PP: 79-119.
- Asensio, M., Castro, Y., Asenjo, E. & Botella, J. (2017). Meta-análisis sobre estudios T&T en museos y exposiciones. *Arkeologi Museoa: Los Cuadernos del Arkeologi*, 8, pp. 115-136.
- Asensio, M., Castro, Y., Asenjo, E., Pol, E., Rodríguez, J.A., Paredes, P., Cabrera,, A., Rodríguez, I. & Villar, C. (2013). Cómo aprender disfrutando de la 'Cocina Valeniana': un modelo de evaluación para el diseño de dispositivos de realidad aumentada. En: Cabrera, A., Rodríguez, I. & Villar, C. (Eds.) *La cocina valenciana del Museo Nacional de Artes Decorativas. Una relectura a través de la tecnología de Realidad Aumentada*. Madrid: Secretaria de Estado de Cultura.
- Asensio, M. & Pol, E. (2017). The Never-ending Story About Heritage and Museums: Four Discursive Models. In: Carretero, M., Berger, S. & Grever, M. (Eds.) *Handbook of Research in Historical Culture and Education*. Capítulo 39. (pp. 755-780). N.Y.: Palsgrave.
- Asensio, M., Santacana, J. & Fontal, O. (2017). Inclusión en Patrimonio y Museos: 'Más allá de la dignidad y la accesibilidad'. *Monográfico de la Revista Hermus*, 17, pp. 39-56.
- Boyle, E.A., Hainey, T., Connolly, T.M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C. & Pereira, J. (2016) An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education* 94 (2016) 178e192.
- Bruce, B.C. (2009). Ubiquitous learning, ubiquitous computing, and lived experience. En B. Cope y M. Kalantzis (Eds.), *Ubiquitous Learning*, (pp. 29-39). Urbana, IL: University of Illinois Press.
- Buckler, T. (2012). Is there an app for that? Developing an evaluation rubric for apps for use with adults with special needs? *The Journal of BSN Honors Research*, 5(1), 19-32.
- Camps-Ortueta, I., Rodríguez-Muñoz, J.M., Gómez-Martín, P.P. & González-Calero, P.A. (2017) Combining augmented reality with real maps to promote social interaction in treasure hunts. In: Camacho, D., Gómez, Y. & González, P.A. *CoSECiVi 2017, IV Congreso de la Sociedad Española para las Ciencias del*

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueología & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

Videojuego. Proceedings of the 4th Congreso de la Sociedad Española para las Ciencias del Videojuego. Barcelona. pp. 131-143.

- Castro, Y., Botella, J. & Asensio, M. (2016). Re-paying Attention: a Meta-Analysis Study of Visitor Behavior. *The Spanish Journal of Psychology*, 19, e39, 1-9.
- Castro, Y., Pérez, C. & Asensio, M. (2017). Construcción de herramientas para el estudio de la inclusión en museología y museografía. *Monográfico de la Revista Hermus*, 17, pp. 77-94.
- Conole, G. & Dyke, M. (2004) Understanding and using technological affordances: a response to Boyle and Cook. *ALT-J, Research in Learning Technology*, 12:3, 301-308. DOI: 10.1080/0968776042000259609.
- Coughlan, J., & Morar, S. S. (2008). Development of a tool for evaluating multimedia for surgical education. *Journal of Surgical Research*, 149(1), 94-100.
- de Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., & Poulouvasilis, A. (2010). Learning as immersive experiences: using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology*, 41, 69-85.
- Earl, M. (2013). iPads in the social classroom, K-12 education. *Instructional Technology Education Special Research papers*, Paper 5.
- Economou, M., and Meintani, E. (2011). Promising beginning? Evaluating museum mobile phone apps. In: *Rethinking Technology in Museums 2011: Emerging experiences*, University of Limerick, Ireland, 26-27 May 2011.
- Falk, J. & Dierking, L. (1992). *The museum experience*. Washington, DC: WhalesBack Books.
- FALK, J. & DIERKING, L.D. (2012): *The Museum Experience Revisited*, Walnut Creek, CA: Sage.
- Csikszentmihalyi, Mihaly (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.
- Csikszentmihalyi, M. (1998). *Finding Flow: The Psychology of Engagement With Everyday Life*. N.Y.: Basic Books.
- Handal, B. (2016). *Mobile makes learning free. Building Conceptual, Professional, and School Capacity*. USA: Information age Publishing Inc.
- Herman, J., & Linn, R. (2013). *On the road to assessing deeper learning: The status of smarter balanced and PARCC assesment consortia, CRESST report 823*. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Hermida, J. (2015). *Facilitating Deep Learning. Pathways to Success for University and College Teachers*. Toronto, New Jersey: Apple Academic Press.
- Kaplan, A. & Patrick, H. (2016). Learning environments and motivation. In: Wentzel, K.R. & Miele, D.B. (Eds.) *Handbook of Motivation at School*. NY: Tylor & Francis. pp. 251-274.
- Lee, C-Y. & Cherner, T. S. (2015). A comprehensive evaluation rubric for assessing instructional apps. *Journal of Information Technology Education: Research*, 14, 21-53. Retrieved from <http://www.jite.org/documents/Vol14/JITEV14ResearchP021-053Yuan0700.pdf>

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

- López-Menchero, V., Grande, A. & Asensio, M. (en prensa). Towards international principles of virtual museums. *Virtual Archaeology Review VAR*.
- Murray, O. T., & Olcese, N. R. (2011). Teaching and learning with iPads, ready or not? *TechTrends*, 55(6), 42-48.
- Norman, D. (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the machine*. N.Y.: Diversion Books.
- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science*, 59(6), 938e955. [SEP]
- Pilgrim, J., Bledsoe, C., & Reily, S. (2012). New technologies in the classroom. *Delta Kappa Gamma Bul- letin*, 78(4).
- Pol, E. & Asensio, M. (2017) *Talking Brains, Resultados de Evaluación*. Barcelona: Fundación La Caixa. Memoria interna no publicada.
- Proctor, N. (2011). De El Museo como Anfiteatro y el Conservador como Anfitrión. En: Asensio, M. y Asenjo, E. (Eds.) *Lazos de Luz Azul: Museos y Tecnologías 1, 2 y 3.0*. Barcelona: UOC. pp. 327-346.
- Rao, L. (2012, January 19). Apple: 20,000 education iPad apps developed; 1.5 million devices in use at schools. *TechCrunch*. Retrieved May 13, 2014, from <http://techcrunch.com/2012/01/19/apple-20000-education-ipad-apps-developed-1-5-million-devices-in-use-at-schools>
- Reeves, T. C., & Harmon, S. W. (1993). Systematic evaluation procedures for interactive multimedia for education and training. *Multimedia Computing: Preparing for the 21st Century*, 472-505.
- Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Englewood, NJ: Educational Technology Publications.
- Sada, A. (2012). *App lists for education*. Retrieved from <http://www.appitic.com/>
- Santacana, J. & Coma, L. (2014) *El M-learning y la educación patrimonial*. Gijón: TREA.
- Santacana, J. & Hernández, X. (2006). *Museología crítica*. Gijón: TREA.
- Santacana, J. & López, V. (2015) *Educación, tecnología digital y patrimonio cultural*. Gijón: TREA.
- Santacana, J. & Llonch, N. (2016). *El patrimonio cultural inmaterial y su didáctica*. Gijón: TREA.
- Santacana, J., Martínez, T. & Asensio, M. (2017). Investigando en Educación Patrimonial: los museos y la inclusión Cultural en la Educación Secundaria. *Monográfico de la Revista Hermus*, 17, pp. 15-24.
- Schibeci, R., Lake, D., Phillips, R., Lowe, K., Cummings, R., & Miller, E. (2008). Evaluating the use of learning objects in Australian and New Zealand schools. *Computers & Education*, 50(1), 271-283.
- Sharples, M., Taylor, J., y Vavoula, G. (2007). A Theory of Learning for the Mobile Age. En R. Andrews y C. Haythornthwaite (Eds.), *The Sage Handbook of Elearning Research*, (pp. 221-47). London: Sage.
- van Boxtel, C., Grever, M. & Klein, S. (Eds.) (2016). *Sensitive Pasts. Questioning Heritage in Education*. NY: Berghahn Books.

Una versió de este paper, se publicó en catalán en: Asensio, M., Santacana J. & Asenjo, E. (2018). Avaluació d'APP per a un patrimoni inclusiu. En: Santacana, J., Asensio, M., & Llonch, N. (Eds.) *App, arqueologia & m-learning. Reconstruir, restituir, interpretar i avaluar app*. Barcelona: Dalmau. Pp. 119-155.

Walker, H. (2010). Evaluating the effectiveness of apps for mobile devices. *Journal of Special Education Technology*, 26(4), 59-66.

Winslow, J., Dickerson, J., & Lee, C. (2013). *Applied Technologies for Teachers*. Dubuque, IA: Kendall Hunt.